

RECEIVED
STIC
BIOTECH/ CHEMICALU.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE

ONLINE SEARCH REQUEST FORM

USER

Spector

U.S. SERIAL NUMBER

1993 SEP 27 11:10 AM 54252491

ART UNIT

1812

PHONE

308-1793

DATE

9/27/95

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and/or relevant claim(s).

*Please search Seq ID No: 1, 2,
418*

FILE

FILE

FILE

FILE

FILE

STAFF USE ONLY

COMPLETED

9-28-95

SEARCHER

SILIP X308 4268

ONLINE TIME

20

TOTAL TIME

25

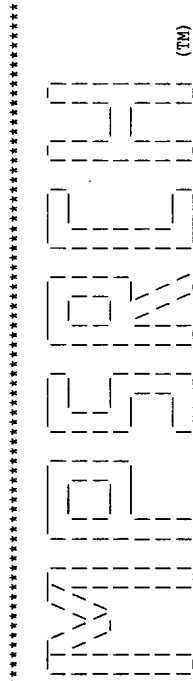
(in minutes)

NO. OF DATABASES

11

SYSTEMS

☐ CAS ONLINE☐ DARC/QUESTEL☐ DIALOG☐ SDC☒ OTHER*MPSEARCH*



```

FT /label= cytoplasmic domain
FT /note= "GC and protein kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 500..602
FT /label= N-glycos_site
PN W09100292-A.
PD 10-JAN-1991.
PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI; 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

DB 2; Score 95; Match 8.9%; QryMatch 6.4%; Pred. No. 1.06e-44;
Matches 88; Conservative 278; Mismatches 615; Indels 9; Gaps 9;

Db 17 vrnrgatntnnavnnhnsyawaarvgnavanavnanvgrannvdrnsvsnngacs 76
:: ||: : : : : ||: : : : : ||: : : : : ||: : : : : ||
Cp 1004 GGGCCAGATGCTGACGGTTCACGGCTCTGAGTGGTTCACAGACAGCCCATGAGT 945
:: ||: : : : : ||: : : : : ||: : : : : ||: : : : : ||

Db 77 nynnnaevdnknyhdnndnngvcvynnaasvarn-ashwrnnnttagaasgnskndh 135
: | :|: : : : : : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 944 TCATTTCACAGGTCGTGTGCTGTCTGATCCAGATTCAGATTCGGGACCTGGA 885
: | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 136 yrtvrtvqneanknngvntvnhqhnwfkaraamnyndardtdrnhvntnngvnanngn 195
: | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 884 GGTGTATTAGTGCACGAGGATATCTTGACTCTGAATCCCTCAAGCCTGCTCAGAAG 825
: | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 196 nsvhnvyarngngnnahtnraagrvcngnnmhmhnnnnnnnnnnnnnnnnnnnnnn 255
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 824 TCCAGGGCCAGGAGTTCTGGCTGTGACACTGAGTTCGTCTCCAAACAATCCAGAGTCT 765
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 256 dvngnsnragtnratgrndtrtnnananrnanntvntvntnnnnnnnnnnnnnnnn 315
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 764 GTTGGGACCTGTTAGTGTGAGGAGTGTGAGACTACTGCTTGGGACAGCTGTGTTGG 705
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 316 rarndngvngnsnnnnn-agcnydgnnnyanvnnntnnnggrndgrnrvnkmgtryh 374

```

```

Cp 704 CAGGGTCCCTCTGACACAGAGGCTGGGACCTTCTACCAAGAGCAGGAAGCGACCTTTC 645
:: ||: : : : : ||: : : : : ||: : : : : ||: : : : : ||
Db 375 gvtgnvmdknnrdntdrnvwangdndsgdnnaahysganknmwftgrnnnwvkgans 434
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 644 CGAAGCAGTGTGTTGCAAGCTCAAGAAGAGGCA-TTGGGGTCTCTGTGAGCTGTGTCTC 586
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 435 dnmncandndsdck-tnnstnanvngtgnthnmngvsnnnnfrkmmnknnaamwr 493
: | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 585 TGGCCCTGTAGAGGAGCTGGCTTCTAGAGAGCCCTGCAGGGCCCCCAAGAGGAGCGGA 526
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 494 nrwnnnngnsrnyhkgagrntnsnrqssysmntahgkykynnmantghknkvvankhv 553
: | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 525 CCTGCCCAAGAGCTGTCCAGAGGAGTGTAGAGGAGGAGGAGGTTTCCAACTGTCTCTG 466
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 554 nkknntnrvnnnkhmrdrnnhnrngacndnnncvntnycnrgenndnnndn 613
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 465 CTGCCATCATCTCCCTCCAGTAGAAGGACACTGCCCTAGAAATGTCTCTGTGCTTGTCT 406
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 614 nndwmnrysnndndvkgmannhnsnshgsknsncvvdsvrvnkntdygnasnrstan 673
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 405 GTTCGCTGTGGGTTTTCATTTCTCCAGGCTAAAGTCCACAGCAGGAGCAGACAGGGA 346
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 674 nddhnanyaknntannnsgnnnttgmaadvysngnnnnnnnnnnnnnnnnnnnn 733
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 345 TAGACAAAGGGTGGAGCTGGGACACTGACTCAGTGGGCTGTGAAGGAGG-TGGGAGTCA 287
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 734 knnvkrngnrynrnsndrtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 793
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 286 CCGACGAGTTTATTAGAGCTCTGGGCTCACAGGACAGCAGCTACGGGCTGCAGAGATT 227
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 794 tgnndnnnnrnnnyannnknnvnrtnaynnnrkanannynnnnnhsvannkrgntvn 853
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 226 A-GCTTGGCCATCGAAGAGCATGGCCGCGAGGAGCAATCAGTCAGCTCCATT-CTGG 169
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 854 anadvntynsdvngvntaneasntnnnnvntnnndnytcndamndndvykvtntngdaym 913
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 168 CCAGGTCTTAAGTGAAGTGGCTCCCTGTATGGGGCTCTCCCTCGAATCT-TTCCTGGG 110
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 914 vvsngnrgnrgnannanmanadavsnrnrhrhdmrnrnrgvhtgnvcagvgnk 973
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 109 GCATGGAGGGGCTTGCAGTCTGTGCTTCTTCGGCAGAGTTCGGGTGGGAGAGGTCG 50
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Db 974 mnrvcmgdvtvntasrmengnanknhvsa 1003
: : :|: : | :|: : | :|: : | :|: : | :|: : | :|: : | :|: : :
Cp 49 -GGCCAAGAGCTCTGTCCGGGTGCAGAGG 21

```

```

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanylyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"

```



```

PT samples
- PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe WK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
- CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

DB 9; Score 43; Match 14.8%; QryMatch 2.9%; Pred. No. 2.56e-11;
Matches 9; Conservative 43; Mismatches 9; Indels 0; Gaps 0;

Db 3 ctcgcgcsghsyvvhvvshhhsvhvhhvhsvvvhhvhhvhhvhhvhyvysvtc 62
||||| : :::::::::: : :::::::::: : :::::::::: : :::::::::: : |||
Qy 1038 CTCAGGGTGGACTTCCTCTCCAAGCCTTGCTCGTAGGACACACACCTTCCTC 1097

Db 63 c 63
|
Qy 1098 C 1098

RESULT 4
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; es.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /*tag= a
FT /function=multiple cloning site
FT primer_bind 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI; 88-27927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diffit base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

DB 1; Score 43; Match 19.3%; QryMatch 2.9%; Pred. No. 2.56e-11;
Matches 32; Conservative 73; Mismatches 58; Indels 3; Gaps 3;

Db 28 cayrcbcagcygcacbyrragg-n-yccccgggywccgagcygaay-ycdchvgcg 85
||| : | : | : ||::: : |||| : ||| : || : : ||:: :

```

[illegible]


```

FT misc_RNA      504..789
-   FT /*tag= f
-   FT /label= mRNA start sequence
-   FT /note= "504 is possible start site"
FT misc_RNA      506..789
-   FT /*tag= g
-   FT /label= mRNA start sequence
FT FT /note= "506 is possible start site"
FT CDS           790..3498
FT FT /*tag= h
FT /label= HSV-1 gB
FT /note= "includes N-terminal hydrophobic leader and
FT a membrane-spanning sequence, a C-terminal
FT ionic sequence, and 9 N-linked
FT saccharide-addition sites"
FT 3'UTR         3499..3549
FT FT /*tag= i
FT polyA_signal  3518..3525
FT FT /*tag= j
FT polyA_site    3549..3549
FT FT /*tag= k
FT misc_RNA      3549..3997
FT FT /*tag= l
FT /note= "3' nonessential sequences to the BamHI
FT site"
PN US4642333-A.
PD 10-FEB-1987.
PF 20-JUN-1984; 622496.
PR 16-SEP-1983; US-532996.
PR 20-JUN-1985; US-622496.
PA (PERS/) PERSON S.
PI Person S;
DR WP; 87-056354/08.
PT Amino acid chain of glycoprotein B of HSV-1 and 2 - prepd. as
PT recombinant and used for vaccines for herpes simplex virus types 1
PT and 2.
PS Example; Table 1; 16pp; English.
CC !NOTE! This sequence has been indexed as represented in the
CC specification, except that bases 'E' have been replaced by 'N'.
CC The features have been indexed according to the legend of table 1 on
CC column 19/20 and the Sequence Summary of column 5 (sic). Note that
CC the base numbering of the features does not correspond to the
CC the sequence numbering below.
CC For another DNA sequence of HS8-1 gB see N71303 (P71135),,
CC and for HSV-2 gB see N71399 (P71136).
CC A pure non-glycosylated amino acid (AA) chain comprising a sequence
CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
CC virus which is antigenic to HSV-1 of HSV-2, which contains no more
CC than 750 AA residues, and which includes AA residues 135-649
CC inclusive is claimed. It can be used to produce vaccines for
CC prophylaxis and treatment of HSV-1 and HSV-2.
SQ Sequence 3871 bp; 743 A; 1402 C; 544 G; 754 T;

DB 2; Score 25; Match 39.4%; OryMatch 1.7%; Pred. No. 1.72e-01;
Matches 54; Conservative 32; Mismatches 49; Indels 2; Gaps 2;

Db 2936 cbtbcggtatcbacbtgcggbbttbctccttcattcatbtcacaacctttctttbbb 2995
| : : | | | : : | : : | : : | : : | : : | : : | : : | : : | : :
Qy 1022 CTCCTGGCATTCAACTCGAGGTGCATTCCTCTCTTCTCAAGCCTTGCTGCTCATGG 1081
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :

Db 2996 bgcbtgcbcbtbbbtcttbcttcbbcgccctggcgbccctcttcobcctttcbtta 3055
: : ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy 1082 ACACACACCCTCCCCTC-CTTCA-CCTGCCCTTGCACACACACCATGGATTCACACCCAG 1139

```

D_b 3056 cbtcatcbbbbtctbc 3072
I: |||::||:: |:
Q_y 1140 CTCACGCCCTGTTTC 1156

RESULT 10
ID Q20695 standard; cDNA; 1542 BP.
AC Q20695;
DT 24-APR-1992 (first entry)
DE Mouse OSF.1.
KW MK; calvarial cell; probe; diagnosis; osteoblast;
KW cranial nerve cell; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 229..735
FT /*tag= a
FT /label= I-I
PN W09200324-A.
PD PD 09-JAN-1992.
PF 27-JUN-1991; J00871..
PR 29-JUN-1990; JP-169824.
PR 28-SEP-1990; JP-256810.
PA (FAHR) HOECHST JAPAN LTD.
PI Hashimoto T, Tezuka K, Kumegawa S, Takagi C;
DR WPI; 92-041516/05.
DR P-PSDB; R20238.
PT Protein from mouse calvarial cells - differentiates growth of
PT osteoblast(s) and cranial nerve cells, for treating and
PT diagnosing osteoporosis and dementia
PS Disclosure; Fig 3(1-5); 43pp; Japanese.
CC A cDNA library was constructed from mouse calvarial cell line
CC MC3T3E1. Differential screening in NIH3T3 cells (ATCC CRL-1658) and
CC cloning, then insertion into vector pUC18 gave plasmid pMC031 which
CC contains the DNA coding for mouse OSF-1 (Q20695). This plasmid was
CC used as a probe to clone by plaque hybridisation from a human cDNA
CC library to give a phage clone, HBRI. Insertion into vector pUC18
CC gave pHBRI which may be used diagnostically or the OSF-1 cDNA
CC (Q20696) incorporated (with poly-A signal sequence and SV40 T-cell
CC antigen promoter) into a vector (such as pHSG-757) for expression
CC (e.g. in CHO cells) of the protein.
CC See also Q20695-97.
SQ Sequence 1542 BP; 488 A; 277 C; 360 G; 417 T;

D_B 3; Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 5.18e-01;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

D_b 149 ccttgagatgcgaacactcgccgcaattgtcaacaagaaggcagtcagtca 204
||||| | |||| | ||||| | || ||||| ||| | |||
Q_y 980 CTGGAGCCTCAGCATCTCGCCCGAGCTTTCACAAGAAGCTCCCTGCATTCA 1035

RESULT 11
ID Q11195 standard; DNA; 36 BP.
AC Q11195;
DT 05-JUN-1991 (first entry)
DE Ballast Constituent coding sequence #2.
KW Ballast constituent; fusion protein; oligonucleotide library; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT repeat unit 7..9
FT /*tag= a
FT /note= "can be present 4 to 8 times"
PN W09103550-A.
PD 21-MAR-1991.

PF 28-AUG-1990; U04840.
 PR 29-AUG-1989; US-399874.
 PA (FARH) HOECHST AG.
 PA (GENO-) GEN HOSPITAL CORP.
 PI Stengelin S, Ulmer W, Habermann P, Uhlmann E, Seed B;
 DR WPT; 91-102070/14.
 PT Prepn. of fusion proteins contg. ballast constituent and protein
 PT - giving prods. which are protease resistant or insoluble
 PS Claim 11; Page 50; 60pp; English.
 CC This oligonucleotide is an example of a member of an oligonucleotide
 CC library encoding ballast constituents. The oligonucleotides are
 CC inserted into a vector, functionally linked to a regulatory region
 CC and to a structural gene encoding a desired protein, e.g. proinsulin.
 CC Host cells transformed with such plasmids produce fusion proteins in
 CC high yield. The ballast is short and does not disturb the folding of
 CC the desired protein. The fusion protein is soluble or easily
 CC solubilised. The oligonucleotide encodes a cleavage site at its 3'
 CC end which allows easy removal of the ballast constituent.
 CC See also Q11194 and Q1196-Q11202.
 SQ Sequence 36 BP; 1 A; 11 C; 3 G; 1 T;

DB 2; Score 23; Match 27.6%; QryWatch 1.5%; Pred. No. 1.52e+00;
 Matches 8; Conservative 18; Mismatches 3; Indels 0; Gaps 0;

[illegible]

RESULT	12	
ID	Q11195	standard; DNA; 36 BP.
AC	Q11195;	
DT	05-JUN-1991	(first entry)
DE	Ballast Constituent coding sequence #2.	
KW	Ballast constituent; fusion protein; oligonucleotide library; ss.	
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	repeat_unit	7..9
FT	/tags= a	
FT	/note= "can be present 4 to 8 times"	
PN	W09103550-A.	
PD	21-MAR-1991.	
PF	28-AUG-1990; U04840.	
PR	29-AUG-1989; US-399874.	
PA	(FARH) HOECHST AG.	
PA	(GEOH-) GEN HOSPITAL CORP.	
PI	Stengelin S, Ulmer W, Habermann P, Uhlmann E, Seed B;	
DR	WPI; 91-102070/14.	
PT	Prepn. of fusion proteins contg. ballast constituent and protein	
PT	- giving prods. which are protease resistant or insoluble	
PS	Claim 11; Page 50; 60pp; English.	
CC	This oligonucleotide is an example of a member of an oligonucleotide	
CC	library encoding ballast constituents. The oligonucleotides are	
CC	inserted into a vector, functionally linked to a regulatory region	
CC	and to a structural gene encoding a desired protein, e.g. proinsulin.	
CC	Host cells transformed with such plasmids produce fusion proteins in	
CC	high yield. The ballast is short and does not disturb the folding of	
CC	the desired protein. The fusion protein is soluble or easily	
CC	solubilised. The oligonucleotide encodes a cleavage site at its 3'	
CC	end which allows easy removal of the ballast constituent.	
CC	See also Q11194 and Q11196-Q11202.	
SQ	Sequence	36 BP; 1 A; 3 G; 1 T;

DB 2; Score 23; Match 24.1%; QryMatch 1.5%; Pred. No. 1.52e+00;

Matches 7; **Conservative** 19; **Mismatches** 3; **Indels** 0; **Gaps** 0;

[illegible]

RESULT 13
ID Q11197 standard; DNA, 33 BP.
AC Q11197;
DT 05-JUN-1991 (first entry)
DE Ballast Constituent coding sequence #4.
KW Ballast constituent; fusion protein; oligonucleotide library; ss.

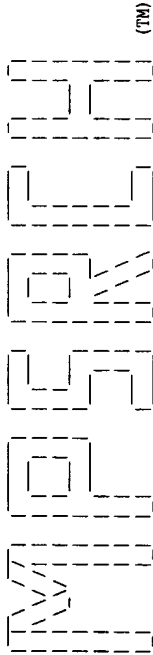
FD	Key	Location/Qualifiers
FT	repeat_unit	7..9
FT	/*tag_ a	
FT	/note= "can be present 4 to 7 times"	
PN	W0910350-A.	
PD	21-MAR-1991.	
PR	28-AUG-1990; U04840.	
PR	29-AUG-1989; US-399874.	
PA	(FARH) HOECHST AG.	
PA	(GEHO-) GEN HOSPITAL CORP.	
PI	Stengelin S, Ulmer W, Habermann P, Uhlmann E, Seed B;	
DR	WPI; 91-102070/14.	
PT	Prepn. of fusion proteins contg. ballast constituent and protein	
PT	- giving prods. which are protease resistant or insoluble	
PS	Claim 12; Page 50; 60pp; English.	
CC	This oligonucleotide is an example of a member of an oligonucleotide	
CC	library encoding ballast constituents. The oligonucleotides are	
CC	inserted into a vector, functionally linked to a regulatory region	
CC	and to a structural gene encoding a desired protein, e.g. proinsulin.	
CC	Host cells transformed with such plasmids produce fusion proteins in	
CC	high yield. The ballast is short and does not disturb the folding of	
CC	the desired protein. The fusion protein is soluble or easily	
CC	solubilised. The oligonucleotide encodes a cleavage site at its 3'	
CC	end which allows easy removal of the ballast constituent.	
CC	See also Q11194-6 and Q11198-Q11202.	
SQ	Sequence 33 BP; 2 A; 10 C; 3 G; 1 T;	

DB 2; Score 22; Match 32.1%; QryMatch 1.5%; Pred. No. 4.35e+00;
Matches 9; Conservative 16; Mismatches 3; Indels 0; Gaps 0;

Db 4 g c a d c d d d d d d d d d d d d d d d d d c d c 31
 ||| : | : | : | : | : | : | : | : | :
Qv 629 G C A C A A C T G C T T C G G G G A A G G T C C G C 656

RESULT	14	
ID	Q46603	standard; DNA; 33 BP.
AC	Q46603;	
DT	20-DEC-1993	(first entry)
DE	Mixed oligonucleotide #5	encodes ballast constituent.
DE	Mixed oligonucleotide #5	encodes ballast constituent; pro-insulin production;
KW	Fusion protein;	ballast constituent; HMG CoA reductase;
KW	recombinant protein production;	HMG CoA reductase;
KW	human 3-hydroxy-3-methylglutaryl-coenzyme A-reductase;	
KW	mixed oligonucleotide; ss.	
OS	Synthetic.	
Key	Location/Qualifiers	
FT	repeat unit	7..9
FT	/*tag=	a
FT	/rpt_type=	tandem
FT	/note=	"can be repeated 4-8 times"

! *****



Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPerch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 27 23:49:23 1995; MasPar time 782.94 Seconds
1107.806 Million cell updates/sec
Tabular output not generated.

Title: >US-08-252-491-1
Description: (1:1486) from US08252491.seq
Perfect Score: 1486
N.A. Sequence: 1 CCTCGTGGCGGTCTGTAGGC.....TATAAATTGAAAATCACTA 1486
Comp: GGAGCAGGCCAGGACTCCG.....ATATTTAAACTTTTAGTGAT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 267821 seqs, 291840210 bases x 2

Database: emb1-new6

- 1 BCT
- 2 EST1
- 3 EST2
- 4 EST3
- 5 EST4
- 6 EST5
- 7 FUN
- 8 INV1
- 9 INV2
- 10 MAM
- 11 ORG
- 12 PLN
- 13 PRI
- 14 PRO
- 15 ROD
- 16 STS
- 17 SYN
- 18 UNC
- 19 VRT
- 20 VIR

Database: genbank89
21 BCT1
22 BCT2
23 BCT3
24 BCT4
25 BCT5

- 26 BCT6
- 27 INV1
- 28 INV2
- 29 INV3
- 30 INV4
- 31 INV5
- 32 MAM1
- 33 MAM2
- 34 PAT1
- 35 PAT2
- 36 PHG
- 37 PLN1
- 38 PLN2
- 39 PLN3
- 40 PLN4
- 41 PLN5
- 42 PLN6
- 43 PLN7
- 44 PRI1
- 45 PRI2
- 46 PRI3
- 47 PRI4
- 48 PRI5
- 49 PRI6
- 50 PRI7
- 51 PRI8
- 52 PRI9
- 53 ROD1
- 54 ROD2
- 55 ROD3
- 56 ROD4
- 57 ROD5
- 58 ROD6
- 59 ROD7
- 60 STR
- 61 STS1
- 62 STS2
- 63 STS3
- 64 STS4
- 65 SYN
- 66 UNA
- 67 VRL1
- 68 VRL2
- 69 VRL3
- 70 VRL4
- 71 VRL5
- 72 VRL6
- 73 VRT1
- 74 VRT2
- 75 VRT3

Database: genbank-new6

- 76 BCT
- 77 EST1
- 78 EST2
- 79 EST3
- 80 EST4
- 81 EST5
- 82 EST6
- 83 EST7
- 84 EST8
- 85 INV
- 86 MAM
- 87 PHG

88 PLN
89 PRI
90 ROD
91 STR
92 STR
93 SYN
94 UNA
95 VRL
96 VRT

Database: u-emb143 89
97 ALL

Statistics: Mean 11.614; Variance 3.738; scale 3.107

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1486	100.0	1486	57	MUSTHROA	0.00e+00
2	676	45.5	1795	50	HUMMLCPL	0.00e+00
3	616	41.5	1341	47	HSU11025	0.00e+00
4	602	40.5	1062	51	HUMTHROMB	0.00e+00
5	328	22.1	7666	51	HUMTA	0.00e+00
6	328	22.1	6163	51	HUMTHROWA	0.00e+00
7	26	1.7	3379	41	SCILVS	4.55e-03
8	25	1.7	335	44	HSDNASDAZ	2.64e-02
9	24	1.6	1525	59	S52337S1	1.46e-01
10	24	1.6	134226	71	IHICG	1.46e-01
11	24	1.6	3940	31	TRECRE1	1.46e-01
12	24	1.6	1514	57	MUSOSF1	1.46e-01
13	23	1.5	5322	50	HUMPKD1GEN	7.59e-01
14	23	1.5	329	67	ADRHUMJ	7.59e-01
15	23	1.5	5322	73	CHKOVALSA	7.59e-01
16	23	1.5	5322	89	HUMPKD1GEN	7.59e-01
17	23	1.5	675	75	S82572S1	7.59e-01
18	23	1.5	350	44	HSDINURPA	7.59e-01
19	23	1.5	875	62	G05108	7.59e-01
20	23	1.5	5322	13	HSPKD1GEN	7.59e-01
21	23	1.5	38139	85	CELF45E1	7.59e-01
22	23	1.5	38139	8	CELF45E1	7.59e-01
23	23	1.5	84539	44	HSABLGR3	7.59e-01
24	22	1.5	1468	48	HUMC2PHTYR	7.59e-01
25	22	1.5	426	6	HS971139	3.71e+00
26	22	1.5	5174	44	HSCOL4A3	3.71e+00
27	22	1.5	2866	37	ANUAPA	3.71e+00
28	22	1.5	1521	48	HUMC1PHTYR	3.71e+00
29	22	1.5	417	6	HS941138	3.71e+00
30	22	1.5	970	47	HUMAAPA	3.71e+00
31	22	1.5	2246	48	HUMCD203	3.71e+00
32	22	1.5	1260	65	SYNBCTCONS	3.71e+00
33	22	1.5	8647	52	S45332	3.71e+00
34	22	1.5	1475	52	S62885	3.71e+00
35	22	1.5	3505	32	BTU03420	3.71e+00
36	22	1.5	467	4	HS586134	3.71e+00
37	22	1.5	970	52	S62884	3.71e+00
38	22	1.5	467	78	R69586	3.71e+00
39	22	1.5	1475	47	HUMAPB	3.71e+00
40	22	1.5	885	55	MUSCSCCE	3.71e+00

41 22 1.5 300 56 MUSINT22 Mouse integration sit 3.71e+00
42 22 1.5 1722 52 S57235 CD68=110kda transmem 3.71e+00
c 43 22 1.5 600 64 HUMUT6643 Human STS UT6643. 3.71e+00
c 44 22 1.5 14977 45 HSFURIN Human fur gene, exons 3.71e+00
c 45 22 1.5 417 80 R75941 yi62a02.r1 Homo sapie 3.71e+00

ALIGNMENTS

RESULT 1	MUSTHROA	1486 bp	mRNA	ROD	06-JUL-1994
LOCUS	Mus musculus thrombopoietin mRNA, complete cds.				
DEFINITION	L34169				
KEYWORDS	thrombopoietin.				
SOURCE	Mus musculus cDNA to mRNA.				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 1486)				
AUTHORS	Lok, S., Kaushansky, K., Holly, R.D., Kuijper, J.L., Lofton-Day, C.E., Oort, P.J., Grant, F.J., Heipel, M.D., Burkhead, S.K., Kramer, J.M., Bell, L.A.N., Sprecher, C.A., Blumberg, H., Johnson, R., Prunkard, D., Ching, A.F.T., Mathewes, S.L., Bailey, M.C., Forstrom, J.W., Buddle, M.M., Osborne, S.G., Evans, S.J., Sheppard, P.O., Presnell, S.R., O'Hara, P.J., Hagen, F.S., Roth, G.J. and Foster, D.C.				
TITLE	Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo				
JOURNAL	Nature 369, 565-568 (1994)				
MEDLINE	94261207				
COMMENT	NCBI gi: 508540				
FEATURES	Location/Qualifiers				
source	1..1486				
CDs	/organism="Mus musculus" /sequenced_mol="cDNA to mRNA" 174..1244 /note="NCBI gi: 508541" /codon_start=1 /function="c-mpl receptor ligand" /evidence=experimental /product="thrombopoietin" /translation="MEITDILLLAAILAVARLTSSVPACDPRLINKLRDLSLLH SRLSQCPDVPDLSIPVLPVAVDSIGEMKTOEQSAQDILGAVSLLEGVMAARGOL EPSCILSLGQLSGQVRLILGALGCLLTQPLQGRTHAKDPNALFLSIQQLLRKV RFLLLVEGPTLCVRLTPTTAVPSSTSSQLLNKFPNRTSGLTETNF SVTATAGPOL LSRLQGRVKITPQNLNQTSPVQISGLYNRTGPNVNTGFCAGTSLQTLASDLS PGAFNKGSIAPNLQGLPPSPSLAPDGHPTTFPPSPALPTTHGSPPLHLFPDPSTTM PNSTAPHPVTWYMPHNPUSQET"				
BASE COUNT	331 a 484 c 337 g 334 t				
ORIGIN					
DB 57;	Score 1486; Match 100.0%; QryMatch 100.0%; Pred. No. 0.00e+00;				
Matches	1486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	1 cctcgtccggtcctgagcccttctccaccggagacagagctcttggcccaactctctcc 60 				
Qy	1 CCTGTCGGCTCTGAGGCCCTTCTCCACCGGAGAGAGTCTTGGCCGACCTCTCC 60 				
Db	61 caccgactctgccaaagacacagaagctcaagccgctccatgccccaggaaga 120 				
Qy	61 CACCCGACTCTCCGAAAGACACAGAGCTCAAGCCGCTCCATGCCCCAGGAAGA 120 				
Db	121 ttccagggagagggcccatcacaggaggacacttcaagtagacacacccagaatggagc 180 				
Qy	121 TTCAGGGAGAGGCCCATACAGGAGGAGCCACTTCAGTTAGACACCTGGCCAGATGGAGC 180 				

Db 181 tgaactgattgtctcctgagccgcatgctcttctgagtggaagcaactaactctgacagcc 240
 Qy 181 TGACTGATTGTCTCTGCGGGCATGCTCTTCGAGTGGCAAGACTTACTCTGTCCAGCC 240
 Db 241 ccgtagctcctgctgtgacccagactcctaaataaactgctgggtgaetcccactcc 300
 Qy 241 CCGTAGCTCTGCTGTGACCCACAGCTCTCTAAATAAATGCTGCGGTGACTCCACCTCC 300
 Db 301 ttacagccgactgagtcagtgctcccaagctcgaccccttctatccctgttctgctgc 360
 Qy 301 TTACAGCCGACTGAGTCAAGTGTCCGACGCTGACCCCTTTGTCTATCCCTGCTTCTGCTGC 360
 Db 361 ctgctgagcatttgcctggagagtggaataaccagagcaagagcaagcaagg 420
 Qy 361 CTGCTGTGAGCTTTAGCTGGGAGATGGAAACCCAGACGGAACAGAGCAAGGACAGG 420
 Db 421 acattctagggcagtgctccttctactgagggagtgatggcagcagagacagtgg 480
 Qy 421 ACATTCTAGGGGCAGTGCTCCTTCTACTGGAGGGAGTGTGGCAGCAGGACAGTGG 480
 Db 481 aacctcctgctctcatccctcctggacagctttctgggaggttgcctcctcttgg 540
 Qy 481 AACCTCTCTGCTCTCATCCTCTCTGGGACAGCTTCTTGGGCGAGGTGCGCTCTCTTGG 540
 Db 541 gggccctgaggggctcctcctaggaacccagcttctctacagggcaggacacagctaca 600
 Qy 541 GGGCCCTGCAGGGGCTCTGTAGGAACCCAGCTTCTCTACAGGGCAGGACACAGCTACA 600
 Db 601 aggaacccaatgccctctcttgagcttgcaacaactgcttcgggaaagggtgcgtcc 660
 Qy 601 AGGACCCCAATGCCCTCTTCTGAGCTTGCAACAACCTCTTGGGGAAGGTGCGCTCC 660
 Db 661 tgctcttgtagaaggtcccaacctctgtgcagaggaacctgccaacacacagctgtcc 720
 Qy 661 TGCTTCTGGTAGAAGTCCACCCCTCTGTCTCAGAGGAGCCCTGCCAACACACAGCTGCC 720
 Db 721 caagcagtaacttctcaactctcacaactaaacaagtctccaaacaggactcttgattgt 780
 Qy 721 CAAGCAGTACTTCTCACTCTCTCACACTAACACAGTTCCTCCAAACGACTCTTGATGT 780
 Db 781 tggagcagaacttcagtcacagccaagaactgctggccctggaactctgagcaggttc 840
 Qy 781 TGGACAGCAACTTCAGTGTCCAGCCAGAACTGCTGGCCCTGGACTTCTGAGCAGGCTTC 840
 Db 841 agggattcagagtcagattactcctggtcagctaaatacaaacctccaggtcccaagtc 900
 Qy 841 AGGGATTCAAGTCAAGATTACTCTCTGCTCAGTAAATCAAACTCCAGGTCGCCAGTCC 900
 Db 901 aaatctctgatacctgaacagagcacacaggaacctgtgaatggaactcatggtctttg 960
 Qy 901 AAATCTCTGGATACCTGAACAGCACACAGGACCTGTGAATGGAATCATGGGCTTTTG 960
 Db 961 ctggaacctcaacttcagacctggaagcctcagacatctcgccggagctttcaacaag 1020
 Qy 961 CTGGAACTCACTTCAGACCTGGAGGCTCAGACATCTGCGCCGAGCTTCAACAAG 1020
 Db 1021 gctcctggcattcaacctccaggttggaacttctcctctctccaaacctgctcctgatg 1080
 Qy 1021 GCTCCTGGCACTCAACCTCCAGGGTGAAGTCTCTCTCTTCTCCAAAGCCTTGCTCTGATG 1080
 Db 1081 gacacacaccccttccctcctcaactgcttgccacacacacacacacacacacacac 1140
 Qy 1081 CACACACACCCCTTCCCTCTACCTGCTTGGCCACCCATGATGATCCACCCAGC 1140

Db 1141 tcaacccctgttctcctgacccctccaccacatgcctaaactctaccgcccctcatccag 1200
 Qy 1141 TCACCCCTGTCTGACCTTCCTGACCTTCACACCATGCTTACTCTTACCGCCCTCATCCAG 1200
 Db 1201 tcaaatgtaccctcatccaggaattgtctcaggaatacatagcggcgacactggccca 1260
 Qy 1201 TCACATGTACCCTCATCCAGGAATTTGTCTCAGGAACATAGCGGGGACTGGCCCA 1260
 Db 1261 gtgagcctctcgacttctctcggggacacagcttccccaggagcctgagagcagctgc 1320
 Qy 1261 GTGAGCCTCTGCACTTCTCTCGGGGCAAGCTTCCCAGGAAGGCTGAGAGGCACTGC 1320
 Db 1321 atctgtccagatgtctcgtcttccataaaggccctgggaaaggatcacacagcactg 1380
 Qy 1321 ATCTGCTCCAGATGTCTGCTTTCACCTAAAGGCCCTGGGAAGGGATACACAGCACTG 1380
 Db 1381 gagattgaaaaatttagagagctattttttaacctatcagaatattcatcagagca 1440
 Qy 1381 GAGATTGTAATTTTAGGAGCTATTTTTTTTAACTATCAGCAATATTTCATCAGACCA 1440
 Db 1441 gctagcagctcttctcgtatcttctcgtataaattgaaatcaacta 1486
 Qy 1441 GCTAGCACTTCTGCTATTTCTGGGTATTAATTTGAAATCACTA 1486

RESULT 2
 LOCUS HUM1.CMPL 1795 bp mRNA PRI 07-JAN-1995
 DEFINITION Human c-mpl ligand (ML) mRNA, complete cds.
 ACCESSION L33410
 KEYWORDS c-mpl ligand.
 SOURCE Homo sapiens cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
 REFERENCE 1 (bases 1 to 1795)
 AUTHORS de Sauvage, F.J., Hass, P.E., Spencer, S.D., Malloy, B.E., Gurney, A.L.,
 Spencer, S.A., Darbonne, W.C., Henzel, W.J., Wong, S.C., Kuang, W.-J.,
 Oles, K.J., Hultgren, B., Solberg, L.A.Jr., Goeddel, D.V. and
 Eaton, D.L.
 TITLE Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl
 ligand [see comment]
 JOURNAL Nature 369 (6481), 533-538 (1994)
 MEDLINE 94261202
 COMMENT NCBI gi: 506826
 FEATURES
 source
 1..1795
 /organism="Homo sapiens"
 /sequenced_mol="cDNA to mRNA"
 216..278
 /gene="ML"
 /codon_start=1
 216..1277
 /gene="ML"
 /note="erythropoietin homology domain bp 66..522; NCBI
 gi: 506827"
 /codon_start=1
 /product="c-mpl ligand"
 /translation="MELTELLIVVLLLTARLTLSPPACDLRVLSKLRLDSVHLH
 SRUSQCEVHPDPTVPLIPAVDFSLGEMTQMEETRAQDILGAVTLLEGVNAARGL
 GPTLSSLLGQSLGQVLLGALQLLGTPLPQGRTHAKDPNAILFLSFQHLIRKQV
 RFLMLVGGSTLCVRRAPPTTAVPSRTSLVLTINELPNRTSLGLETNFTASRTGCSGL
 LKMQQGRAKIPGLIQTSLDQIPGYLNRIHELINGTRGLFFGPSRRTLGADPISL
 GTSDTGSLPPNIQPGYSPSTPPTGQVTLFFLPPTLPTPVVQLHPLLPDPSAPTTP

```

mat_peptide      TSPLLNTSTHSONLSOEG*
                  279..1274
                  /gene="ML"
                  /codon_start=1
                  /product="c-mpl ligand"
BASE COUNT      420 a 586 c 382 g 407 t
ORIGIN
DB 50; Score      676; Match 80.7%; QryMatch 45.5%; Pred. No. 0.00e+00;
Matches 967; Conservative 0; Mismatches 219; Indels 12; Gaps 7;
Db 85 cettggccgcctttggcccccactactctgccagaagtga-agagcct--aagccgcc 142
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 42 CCTTGGCCGACCTCTCCACCCGACCTGCGCGAAGACAGACAGAGCTCAGCGCGCC 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 tccatggcccccaggaagattcaggggagagcccaaacagagagccacgagccaga 202
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 102 TCCATGGCCCCAGGAAAGATTACGGGGAGAGGCCCCCATACAGGGGAGCCACTTCAGTTAGA 161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203 caccggccagagaatggagctgactgaatgctcctctgtgtcatgctctcctaactgc 262
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 162 CACCT-GGCCAGAAATGAGAGTGACTGATTTGCTGCTCGCGGCCATGCTTCTTGCACTGGC 220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 aaggctaacgctgtccagccgggctcctcctgtctgtgaactccgagctcagtaaaact 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 221 AAGACTAACTCTGTCCAGCCCGCTGCTCTGCTGTGACCCGACGACTCTTAATAAATACT 280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 gcttgactgactccttccacagagactgagccagtgccagaggttcacccctt 382
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 281 GCTGCTGACTTCCACCTCTCTACAGCCGACGTGAGTCACTGTCCGAGAGTCCGACCTTT 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 gctacacctctgctgctgctgtgactttagcttgggagaatggaaaaccagat 442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 341 GCTATCCTCTGTTGCTGCTGCTGTGAGCTTTAGCCCTGGGAGATGGAAACCCAGAC 400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 443 ggaggagaccagagcacagacattctggagagcagtgaacctctcgtggaggagtgat 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 401 GGAACAGAGCAAGGACAGACGATTTAGGGGAGAGTGTCCCTCTTACTGGAGGAGTGAT 460
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 ggcagcaggggacaaactgggaacacacttgcctctcatccctcctggggcagcttctcg 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 461 GGCAGCAGGAGGACAGTTGGAAACCTCTGCTCTCATCCCTCTCTGGCAGAGCTTCTGG 520
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 563 acagtcgctctcctcttggggccctgcagagcctccttggaaaccagcttctccaca 622
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 521 GCAGGTTGGCTCTCTTGGGGGCGCTCGAGGGCGCTCTAGGAAACCGAGCTTCTCTACA 580
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 623 gggcagaccacagctcacagagatcccaatgcacatctcctgagcttccaaacacgtct 682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 581 GGGCAGGACACAGCTCACAGGAGCCCAATGCCCTCTTCTTCTGAGTTGCACAACTGCT 640
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 683 ccgagaaaggtgcgtttcctgactgttagagaggtccacacctcgtcagcggggc 742
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 641 TCGGGGAAAGGTGCGCTTCTGCTGTGAGAGGTCCACCCCTCTGTTGTCAGAGCGGAC 700
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 cccaccacacagctgtccccagcagaacctctctagtccacactgaacagctccc 802
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 701 CCTGCTCAACACACAGCTGTCCCAAGCAGTACTTCTCACTCTCTCACACTAAACAAGTTCCC 760
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 803 aaacagactcttgattgttgagacaacttcaactcctcagccagcagaactactggctc 862
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 761 AAACAGGACTTCTGATTTGTGGACGACCACTTCAGTGTACAGGCCACAACTGCTGGCCC 820
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 863 tgggctcttgagtggtgagcagggagattcagagccaagatt---cctggtctgctgaacca 919

```

```

Qy 821 TGGACTTCTGACGAGGCTTCAGGGAATTCAGAGTCAAGATTACTCTGTCAGCTAAATCA 880
      ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 920 aaactccaggtccctggacaaaatccccggataacctgaacagagataacagaactcttga 979
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 881 AACCTCCAGGTCGCCAGTCCAAATCTCTGGATACCTGAACAGGACACACGACCTGTGAA 940
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 980 tgaactcgtggactcttcttggaccttcacgagcagcagcagcagcagcagcagcagcattc 1039
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 941 TGGAACTCATGGGCTCTTCTGGAACCTCACTTCAGACCCCTGGAGCGCTCAGACATCTC 1000
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1040 ctccaggaacatcagacacagctccttgcacccaacctccagcctggatattctcttc 1099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1001 GCGCGGAGCTTCAACAAAGGCTCCCTGGGATTCACCTGCGAGGTGGACTTCTCTCTTC 1060
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1100 cccaaccatcctcactggacagt-a---tacgtctcttccctcttcca---ccacactt 1153
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1061 TCCAGGCTTGTCTCTCATGAGACACACCCCTTCCCTTCACTGCTTGCCTTGCACCCAC 1120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1154 gccacccctctggtccagctccacccctctgcttccagccttctgctccaaagccac 1213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1121 CCATGATCTCCAGCCAGCTCCACCCCTGTTCTGACCTTCCACCAACCATGCTTAA 1180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1214 cctacacgacctttttaaacaatcctacaccactccagaaatctgtctcaggaa 1271
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1181 CTCTACGGCCCTCATCCAGTCACAATGTACCCCTCATCCAGGAATTTGTCTCAGGAA 1238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT          3
LOCUS           HSU11025      1341 bp      mRNA          PRI      15-OCT-1994
DEFINITION      Human megakaryocyte growth and development factor (MGDF) mRNA,
complete cds.
ACCESSION       U11025
KEYWORDS        .
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryotae; Eukaryota; Eukaryote crown group; Metazoa/Eumycota
group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia;
Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopherygii;
Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 1341)
AUTHORS         Bartley,T.D., Bogenberger,J., Hunt,P., Li,Y.S., Lu,H.S., Martin,F.,
Chang,M.S., Samal,B., Nichol,J.L., Swift,S. et.al.
TITLE           Identification and cloning of a megakaryocyte growth and
development factor that is a ligand for the cytokine receptor Mpl
JOURNAL         Cell 77 (7), 1117-1124 (1994)
MEDLINE         94291201
REFERENCE       2 (bases 1 to 1341)
AUTHORS         Samal,B.B.
TITLE           Direct Submission
JOURNAL         Submitted (18-JUN-1994) Babru B. Samal, Developmental Biology,
Angen Inc., Angen-center, Thousand Oaks, CA 91320, USA
COMMENT         NCBI gi: 511223
FEATURES         Location/Qualifiers
source          1..1341
                /clone="HFL-IVC7"
                /organism="Homo sapiens"
                /tissue_type="liver"
                /dev_stage="fetal"
                36..1097
                /gene="MGDF"
                /note="NCBI gi: 558078"
                /codon_start=1
CDS

```

/product="megakaryocyte growth and development factor"
 /translation="MELTELLLVVMLLTARLTSSPAPPACDLRVLSKLDRSHVLH
 SRLSQPEVHPLETPVLLPAVDFSLGEMKQMEETKAQDILGAVTLLEGVMAARGOL
 SPTGLSSLLGELSQVRLILGLALSLGTLQPLPQCRTHAKDPNRAIFLSFQHLRGKV
 RFLMLVGSTLGVRRAPPTTAVPRTSLVLTINELPNRTSGLETFNTASRTTSGSL
 LKWOQGFRAKIPGLINQTSRSLDQIPGYLNRILHLLNGTRGLFPGPSRRTLCAPDISS
 GTSDTGSLLPNLQPGYSPSPTHPTTQGYTLFPLPTLPVTVVQLHPLLPDPSAPPTTP
 TSPLLNTSYTHSQNLSDQG"
 99..1094

mat_peptide

/gene="MGDF"
 /product="megakaryocyte growth and development factor"
 BASE COUNT 298 a 452 c 294 g 297 t
 ORIGIN

DB 47; Score 616; Match 80.3%; QryMatch 41.5%; Pred. No. 0.00e+00;
 Matches 883; Conservative 0; Mismatches 207; Indels 10; Gaps 5;

Db 1 acaggagccagccagccagacacccccggcagaatggagctgaattctctctcg 60
 Qy 140 ACAGGAGCCACTTCAGTTAGACACCT--GGCGAGAATGGAGCTGACTGTTTCTCCTGG 198
 Db 61 tggctgatgtctcttaactgaaggttaacgtgtccagccgggtctctctctgtgtg 120
 Qy 199 CGGCATGCTTCTTCAGCTGGCAAGACTACTCTCTCCAGCCCGCTAGCTCTGCCTGTG 258
 Db 121 acctcaggtctctgaataactgcttgaactccactgctctctccagcagactgagcc 180
 Qy 259 ACCCAGACCTCTTAATAAATGCTGCGTGACTGCCACCTCTCTACAGCGGACTGAGTC 318
 Db 181 agtgcacagaggttcaccccttgctctacacgtgtcctgtgcctgtgcctgtgactttagct 240
 Qy 319 AGTGCCCGACGCTCGACCCCTTGTCTATCCCTGCTTCTGCTGCTGCTGACTTTAGCC 378
 Db 241 tgggagaatggaacccagatggaggagaccagagacaggtatctctgggagcagtga 300
 Qy 379 TGGGAGATGGAAACCCAGACGGAACAGGCAAGGACAGGACATTTAGGGGCACTGT 438
 Db 301 cctcttctgtgaggaggtgatggcagcagcgggagacaactgggaccactgttctctcat 360
 Qy 439 CCTCTTACTTGAGGAGCTGATGGCAGCAGCAGCAGCTTGGAACCTCTCTGCTCTCAT 498
 Db 361 cctctctggggagcttcttgacaggtccgctctcctcttggggcctgcagagcctcc 420
 Qy 499 CCTCTGGGACAGCTTCTTGGGAGGTTGCGCTCTCTTGGGGGCGCTCGAGGGCCTCC 558
 Db 421 ttggaacccaggttctctccagggcagagaccagagctcacaaggatcccaatgccatct 480
 Qy 559 TAGGAACCCAGCTTCTCTCAGGGCAGACACAGCTCACAAGGACCCCAATGCCCTCT 618
 Db 481 tctcagcttccacacactcctccaggaaggtgcgttctctgatgtctgtaggaggtt 540
 Qy 619 TCTTGAGCTTGCAACAACTGCTTGGGGGAAAGTGCGCTCTCTTCTGCTAGGAAGTCT 678
 Db 541 ccacccctctgctcaggcggggcccccacacacacagctgtccccagcagaacctcttag 600
 Qy 679 CCACCCCTCTGTCTCAGACGGACCCCTGCCACACACAGCTGTCCACAGGACTTCTCAAC 738
 Db 601 tctcacaactgaacgagctcccaaacagagactcttgattgtttggagacaaactcactg 660
 Qy 739 TCTCACACTAACAAAGTTCCTCCAAAAGAGACTTCTGGATTTGTCAGACGAAGCTTCAAGT 798
 Db 661 cctcagcagaactactggctctgggctctgaagtggcagcaggggattcagagccaaga 720
 Qy 799 TCACAGCCAGAACTGCTGGCCCTGGACTTCTGAGCAGGCTTCAGGATTTCAGAGTCAAGA 858

Db 721 tt----cctgtgtctgaacaaacctccaggtccctggaccacaaatccccggatacctga 777
 Qy 859 TTACTCTTGGTCAGCTAAATCAAACTCCAGGTCCCAATCTCTGGATACCTGA 918
 Db 778 acagatatacacgaactgaatgaactcgtggaactcttctctggacctcaagcagga 837
 Qy 919 ACAGGACACACGGACCTGTGAATGGAATCAATGGGCTCTTCTCGGAACCTCACTTCAGA 978
 Db 838 cctcagagcccccagacattctccagaaacatcagacacaggtctccctgccaccaacc 897
 Qy 979 CCTCGAAGCTTCAGACATCTCGCCCGGAGCTTTTAAACAAAGGCTCCCTGGCATTTCAACC 1038
 Db 898 tccagctggatattctctctcccaaacccactcctcactctactggacagt-a----taagctc 953
 Qy 1039 TCCAGGTGGACTTCTCCTCTCCAAAGCTTGTCTCTGATGGACACACAGCTTTCCTC 1098
 Db 954 ttccctcttcca--cccaacttgccacccctgtggtccagctccacccctgttctctg 1011
 Qy 1099 CTTCACTGCTTGGCCACACCACTGATCTCCACCCAGCTCCACCCCTGTTTCTCTG 1158
 Db 1012 acccttctgtccaaagcccccctaccagcctcttctaaacacatcctacacccact 1071
 Qy 1159 ACCCTTCCACACCACTGCTTACTCTACGCGCCCTCATTCAGTCACAAATGTACCTCATC 1218
 Db 1072 cccagaatctgtctcaggaa 1091
 Qy 1219 CCAGGAATTTTGTCTCAGGAA 1238

RESULT 4
 LOCUS Human thrombopoietin mRNA 1062 bp mRNA PRI 22-DEC-1994
 DEFINITION Human thrombopoietin mRNA, complete cds.
 ACCESSION L36052
 KEYWORDS thrombopoietin.
 SOURCE Homo sapiens cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
 REFERENCE 1 (bases 1 to 1062)
 AUTHORS Foster,D.C., Sprecher,C.A., Grant,F.J., Kramer,J.M., Kuijper,J.L.,
 Holly,R.D., Whitmore,T.E., Heipel,M.D., Bell,L.A.N., Ching,A.F.,
 McGrane,V., Hart,C., O'Hara,P.J. and Lok,S.
 TITLE Human thrombopoietin: gene structure, cDNA sequence, expression,
 and chromosomal localization
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (26), 13023-13027 (1994)
 MEDLINE 95108091
 COMMENT NCBI gi: 533216
 FEATURES
 source Location/Qualifiers
 1..1062
 /organism="Homo sapiens"
 /sequenced_mol="cDNA to mRNA"
 1..1062
 /note="putative; NCBI gi: 533217"
 /codon_start=1
 /product="thrombopoietin"
 /translation="MELTELLLVVMLLTARLTSSPAPPACDLRVLSKLDRSHVLH
 SRLSQPEVHPLETPVLLPAVDFSLGEMKQMEETKAQDILGAVTLLEGVMAARGOL
 SPTGLSSLLGELSQVRLILGLALSLGTLQPLPQCRTHAKDPNRAIFLSFQHLRGKV
 RFLMLVGSTLGVRRAPPTTAVPRTSLVLTINELPNRTSGLETFNTASRTTSGSL
 LKWOQGFRAKIPGLINQTSRSLDQIPGYLNRILHLLNGTRGLFPGPSRRTLCAPDISS
 GTSDTGSLLPNLQPGYSPSPTHPTTQGYTLFPLPTLPVTVVQLHPLLPDPSAPPTTP
 TSPLLNTSYTHSQNLSDQG"
 BASE COUNT 223 a 375 c 236 g 228 t

ORIGIN

DB 51; Score 602; Match 80.4%; QryMatch 40.5%; Pred. No. 0.00e+00;
Matches 856; Conservative 0; Mismatches 200; Indels 9; Gaps 4;

Db	1	atggagctgactgaattgctctcctcgttgctcatgctctcctaaactgcaagcgctaaacgtg	120
Qy	174	ATGGAGCTGACTGATTTTGTCTCTGGGGCCATGCTTTTCGAGTGGCAAGCACTAACTCTG	233
Db	61	tcaagcccggtcctccctgcttgtagctccgagtcctcagtaaactgcttgtagctcc	120
Qy	234	TCCAGCCCGCTAGCTCCTCGCTGTGACCCAGCACTCCTAAATAAACTGCTGGCTGACTCC	293
Db	121	catgtccttcacagcagactgagccagtgcccagagttcaccccttgcctacaactgctc	180
Qy	294	CACCTCCTTCACAGCGCACTCAGTCAAGTCCCGAGCTCGACCTTTGCTATATCCCTGTT	353
Db	181	ctgctgctcgtgctggacttttagcttggagaaatggaaacccagatggaggagaccaag	240
Qy	354	CTGCTGCCCTGCTGTGACTTTAGCCCTGGGAGATGGAATCCCAACCCAGACGGACAGCAAG	413
Db	241	gcacaggacattctggagcagtgacctctgctggaggaagtgatggcagcacgggga	300
Qy	414	GCACAGGACATTCCTAGGGCGAGTGCCTTTCTACTGGAGGGAGTGTGGCAGCAGCAGGA	473
Db	301	caactgggacccacttgctctcatcctcctgggcagctttctggacaggtccgtctc	360
Qy	474	CAGTTGGAAACCTCTGCTCTATCCCTCTGCGGACAGCTTTTGGGAGAGTGGCCTC	533
Db	361	ctccttggggccctgcagagcctccttggaaacccagcttctccacagggcagagaca	420
Qy	534	CTCTTGGGGGCCCTGCAGGGCCTCTAGAACCCAGCTTCCTCTACAGGCGAGGACCA	593
Db	421	gctcacaaagatcccaatgccaattcctgagagttccaacactgctccgagaaaaggtg	480
Qy	594	GCTCAACAGGACCCCAATGCCCTCTCTTTGAGCTTGCACAACTCTCTCGGGGAAAGTGC	653
Db	481	cgtttctcgtgcttgtgaggagggtccaacctctggctcaggcggggcccacacacaca	540
Qy	654	CGCTTCTCTGCTTGTAGAGGTGCCACCCCTCTGTCTCAGCAGGACCCCTGCCAACACCA	713
Db	541	gctgtccccagcagaacctctctagtctcacactgaacagctcccaacaggacttct	600
Qy	714	GCTGTCCACGAGTAGTTCTCAACTCCTTCACACTAAACAGTTCCCAACAGGACTTCT	773
Db	601	ggattgttgagacaaacttcaactgctccagccagaactaactggctctgggctttgga	660
Qy	774	GGATTGTTGAGAGGACTTCAGTGTACAGCCGAGAACTGCTGGCCCTGACTTCTGAGC	833
Db	661	tggcagcagggattcagagccaaagatt---cctggctctgtgaacaaaacctccaggtcc	717
Qy	834	AGCTTTCAGGATTTCAGAGTTCAGATTACTCTGGTTCAGCTTAATTCAACTTCAGAGTCC	893
Db	718	cttgaccaaaatcccggatacctgaacaggatcacagaaactctgaatggaactcgtgga	777
Qy	894	CAAGTCCAAATCTCTGGATACCTTGACACAGGACAACAGGACCTGTGAATGGAATCATGGG	953
Db	778	ctctttctggacctcacgcagggaacctaggagcccggaacatttctcaggaacatac	837
Qy	954	CTCTTTGCTGGAACCTCACTTCAGACCCCTGGGAAGCCTTCAGACATCTCGCCCGGAGCTTC	1013
Db	838	gacacaggctccctgccacccaacctccaagctggataattctcttccccacccactcct	897
Qy	1014	AACAAGGCTCCCTGGCATTTCAACTCCAGGGTGGACTTCTCTCTCTCCCAAGCCCTGCT	1073

Db	898	ctactggaagt-a----taagctcttctcttcca---cccacttgcacacccctgtg	951
Qy	1074	ctctatggacacacaccccttccctcttcaactgccttgcgccaccacccatggatctcca	1133
Db	952	gtccagctgcacacccctgcttctctgaccccttctgtccaaagccacccctacacagcct	1011
Qy	1134	ccccagctccacccctcttttccaccttccacacatgcttaactctacgccctct	1193
Db	1012	ctctctaacacactctctacacccactcccagaatctgtctcaggaa	1056
Qy	1194	catccagtcacaaatgtacctctatcccaggaaatttctctcaggaa	1238

```

5
RESULT LOCUS HUMTA 7666 bp DNA PRI 03-MAR-1995
DEFINITION Human gene for thrombopoietin.
ACCESSION D32046
KEYWORDS thrombopoietin.
SOURCE Homo sapiens blood DNA, clone lambdaHGT1.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
1 (bases 1 to 7666)
AUTHORS Sohna,Y., Akahori,H., Seki,N., Hori,T., Ogami,K., Kato,T.,
Shimada,Y., Kawanura,K. and Miyazaki,H.
Molecular cloning and chromosomal localization of the human
thrombopoietin gene
JOURNAL FEBS Lett. 353 (1), 57-61 (1994)
MEDLINE 95010765
COMMENT Submitted (08-Jul-1994) to DDBJ by:
Yoshiaki Soma
Pharmaceutical Research Laboratory
Kirin Brewery Co., Ltd.
2-2 Souja-machi 1 chome
Maebashi
Gunma 371
Japan
Phone: 0272-54-8618
Fax: 0272-52-2307.

```

NCBI gi: 577319	Location/Qualifiers
1..7666	
/organism="Homo sapiens"	
/sequenced_mol="DNA"	
/tissue type="blood"	
<1420...1486	
/number=1	
/label=exon1	
3157..3314	
/number=2	
/label=exon2	
join(3302..3314,3546..3673,3960..4046,5979..6146,	
6383..7048)	
/note="NCBI gi: 577320"	
/codon_start=1	
/product="thrombopoietin"	
/translation="METELLVWMLLTARLTSSPAPPACDLRVLSKLRLDSHWLH SRLSQPEVHPITPVLPAVDVSLGSKTQMEETKAQDILGAVTLLIEGVMAARGOL GRCTCLSLGQLSGVRLILGALQSLIGTLPQQGRTAKDPNAIFLSFOHLRGKV REFLWVGSTLCVRRAPPTTAVPSRSTSVILTNELNRTSGLTNETASARTTGSGL LWQQGFRKIPGLINQTSRLDDQIPGYNRLIRLHLLNGTRGLPFGFSRRTLGAPDTS FTSDTSGLPNQLQPGYSPGTHPTQGYTITFFLPPTLPVTVVQLHLLPDSAPTTPT	

Matches 528; Conservative 0; Mismatches 146; Indels 9; Gaps 4;

```
Db 5039 tctctgcctcagcttccctccagggcaggacacagctcacagatcccaatgccca 5098
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 556 TCTAGGACCCAGGCTTCCTTACAGGGCAGGACCACAGCTCACAAGGCCCAATGCC 615

Db 5099 tcttctgaacttcaacacctgctccaggaaggtcgctttcttgatgctgtaggag 5158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 616 TCTTCTGAGCTTGCACAACACTCTCTGGGGGAAGGTCGCTTCTGCTTCTGAGAG 675

Db 5159 ggtcacacctctgctcagggcggggcccccacacacagctgtcccacagacagaccttc 5218
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 676 GTCCACCCCTCTCTGTGACAGGAGCCCTGCCAACCCACAGCTGTCCCAAGCAGTACTTCTC 735

Db 5219 tagtctcacactgaacgagctcccaacaggacttctggattgttgagacaaacttca 5278
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 736 AACTCTCACACTAACACAGTTCACCAACAGGACTTCTGCAATTGTTGCAGACGACTTCA 795

Db 5279 ctgcctcagccagaactactggtctcggtctctgaagtggcagcagggtattcagagcca 5338
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 796 GTGTCAGAGCCAGAACTGCTGGGCCCTTGGACTTCTGAGCAGGCTTCAGGAGTTCAGAGTCA 855

Db 5339 agattt---cctggtctgctgaaccaaaacctccaggtccctggaccaaaatccccggatacc 5395
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 856 AGATTACTCTGCTGCTCAGCTAAATCAAGACTCCAGGTCGCCAGTCCCAAAATCTCTGGATACC 915

Db 5396 tgaacaggatatacagaactcttgaatggaactcgtggactcttctcaggacccctcaagca 5455
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 916 TGAACAGGACACACGACCTGTCAATGGAACCTCATGGCTCTTTGCTGGAACCTCACATTC 975

Db 5456 ggacctaggagcccgagacttctcaggaaatcagacacagagctccctgcacacca 5515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 976 AGACCTGTGAAGCCTCAGACATCTCGCCGGAGCTTTCACAAAGGCTCCCTGGCATTTCA 1035

Db 5516 acctcagctggatattctcttcccacacctcctcctactggacagt-a---tacg 5571
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1036 ACCTCCAGGTGAGTCTTCTCTCTTCACAGCCTTGCTCTGATGGAACACACCTTCC 1095

Db 5572 ctcttcctcttcca---cccacttgccacccctggtgtccagctccacccctgcttc 5629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1096 CTCCTTCACCTGCTTGGCCACACCCATCGATCTCCACCCCGAGCTCCACCCCGCTTTC 1155

Db 5630 ctgacctcttgctccaagcccccacctaccagcctcttctaaacacatcctcaacccc 5689
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1156 CTGACCTTCCACCACTGCTTAACCTCAGCCGCCCTCATCCAGTCAATGTACCCCTC 1215

Db 5690 actccagaatctgtctcagga 5712
||||| ||||| ||||| |||||
Qy 1216 ATCCGAGGATTTGTTCTCAGGA 1238
```

```
RESULT 7
LOCUS SCILV5 3379 bp DNA PIN 12-SEP-1993
DEFINITION Yeast ILV5 gene for acetohydroxyacid reductoisomerase (EC
1.1.1.86).
ACCESSION X04969
KEYWORDS acetohydroxy acid isomeroreductase; ILV5 gene.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
Eumycota; Ascomycotina; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 3379)
AUTHORS Petersen,J.G. and Holmberg,S.
```

TITLE The ILV5 gene of Saccharomyces cerevisiae is highly expressed
JOURNAL Nucleic Acids Res. 14 (24), 9631-9651 (1986)
MEDLINE 87117524
COMMENT Data kindly reviewed (19-JUN-1987) by Holmberg S.

```
FEATURES             Location/Qualifiers
     source            1..3379
                        /organism="Saccharomyces cerevisiae"
     misc_feature       813..826
                        /note="put. leucine control sequence"
     misc_feature       852..857
                        /note="regulatory sequence"
     misc_feature       880..885
                        /note="regulatory sequence"
     promoter           1031..1037
                        /note="TATA-like sequence"
     misc_feature       1052..1057
                        /note="regulatory sequence"
     misc_feature       1111
                        /note="transcription start site"
     CDS                1192..2379
                        /note="acetohydroxyacid reductoisomerase (AA 1-395); NCBI
                        gi: 3827"
                        /codon_start=1
                        /translation="MLRTQMARLIONSRVITAKTEALATRAAAYSRAARFVKPMT
                        TRGKINFGTETVIERADWPREKLIDYFNNDITFALIGYSGQCYGQGLNRDNGLN
                        VIIGKDGASWKAIEDGWPGKNLFTVEDAIKRGSYVMNLLSDAAQSTWPAIKPL
                        LTKGKTLVSHGFSPVFKDLTHVEPPKDLVILVAPKSGRTVRSLFKEGRGINSYA
                        VMDVTGAHEKAQALVAIGSVYVQYTFEREVNSDLYGERCCLAGGHCHELAQYD
                        VLENGHSPSEAFNEETVEEATOSLYPLIGKYGMDYMDACSTTARRGALDWTPIFNNA
                        LKPVFDLYESTKNGTETKRSLEFNSQPDYREKLEKELDTIRNMEIWKYGEVVKLRP
                        ENQ"
     misc_feature       2544..2560
                        /note="termination control sequence"
     misc_feature       2594..2598
                        /note="transcription termination region"
     CDS                3089..>3379
                        /note="open reading frame (AA 1-97); NCBI gi: 3828"
                        /codon_start=1
                        /translation="MSEPAQKQKVANNISLEQIKASGTVVWADTDFGSGIAKFPQDQS
                        TTNPSSLILAAAKOPTYAKLIDVAVEYKKGKGTTEEQVENAVDRLILVEFGKEI"
```

```
BASE COUNT 1020 a 714 c 693 g 952 t
ORIGIN
DB 41; Score 26; Match 88.2%; QryMatch 1.7%; Pred. No. 4.55e-03;
Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 1142 taggagttatattttttaccctaccagcaatat 1175
||||| ||| ||||| ||||| ||||| |||||
Qy 1396 TAGGAGCTATTTTTTTTAACTATCAGCAATAT 1429

RESULT 8
LOCUS HSDNASDZ 335 bp DNA PRI 01-JAN-1995
DEFINITION H.sapiens DNA sequence.
ACCESSION Z22340
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcotrygii; Chonata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
```

REFERENCE	Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 335) Zabarovsky,E., Kashuba,V., Pettersson,B., Petrov,N., Zakhar'yev,V., Gizatullin,R., Lebedeva,T., Bannikov,V., Erlandsson,R., Uhlen,M., Kiselev,L. and Klein,G. Shot-gun sequencing strategy for long range genome mapping: first results Unpublished 2 (bases 1 to 335) Zabarovsky,E. Direct Submission TITLE JOURNAL Submitted (01-APR-1993) to the EMBL/GenBank/DDBJ databases. Zabarovsky E., Karolinska Institute, Tumor Biology, S-104 01 Stockholm, Sweden COMMENT NCBI gi: 608819 FEATURES Location/Qualifiers source 1..335 /organism="Homo sapiens" /clone="NL122060" /cell_line="MCH903.1" /clone_lib="NLI" /chromosome="chromosome 3" /germline BASE COUNT 38 a 149 c 99 g 16 t 33 others ORIGIN DB 44; Score 25; Match 39.3%; QryMatch 1.7%; Pred. No. 2.64e-02; Matches 22; Conservative 17; Mismatches 17; Indels 0; Gaps 0; Db 279 agggdgdgncckeyyyvvccccaanacckmmctttcccvcvmggddttdtg 334 :::: : :: : :::: :: : Cp 685 AGGTGGGACCTTCACAGAACGAGCGCACCTTTCCCGACGAGTGTGTG 630 RESULT 9 LOCUS SS2337S1 1525 bp DNA 28-MAY-1993 DEFINITION OSF-1=pleiotrophin [mice, DBA/2J, liver, Genomic, 1525 nt, segment 1 of 5]. ACCESSION S52337 KEYWORDS . SEGMENT 1 of 5 SOURCE mice liver DBA/2J. ORGANISM Mus sp. Unclassified. REFERENCE 1 (bases 1 to 1525) AUTHORS Katoh,K., Takeshita,S., Sato,M., Ito,T. and Amann,E. TITLE Genomic organization of the mouse OSF-1 gene JOURNAL DNA Cell Biol. 11 (10), 735-743 (1992) MEDLINE 93090271 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibseq 120279] from the original journal article. This sequence comes from Fig. 4. COMMENT Map location: 7q22-qter. FEATURES NCBI gi: 298209 source Location/Qualifiers 1..1525 /organism="Mus sp." /note="mice" BASE COUNT 402 a 350 c 395 g 378 t ORIGIN DB 59; Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 1.46e-01; Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db	1388	ctcgagctctgcagaaacctgcgcgcgaactttgcacaaaggcagcagctagtca	1443
Qy	980	CCTGGAGCCTCAGACATCTCGCGGAGCTTTCAACAAGAGGCTCCTGGCATCA	1035

RESULT	10		
LOCUS	IH1CG	134226 bp	DNA
DEFINITION	Ictalurid herpesvirus 1 (channel catfish virus [CV1]), strain auburn 1, complete genome.		
ACCESSION	M75136		
KEYWORDS	DNA polymerase; SFI helicase; Zn-binding protein; dUTPase; glycoprotein; membrane protein; protein kinase; serine protease; terminase; thymidine kinase.		
SOURCE	Ictalurid herpesvirus 1 (strain auburn 1) (library: ATCC VR-665) DNA.		
ORGANISM	Ictalurid herpesvirus 1		
REFERENCE	Viridae; de-DNA enveloped viruses; Herpesviridae.		
AUTHORS	1 (sites)		
TITLE	Harrison, P. T., Thompson, R., and Davidson, A. J.		
TITLE	Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinase		
JOURNAL	J. Gen. Virol. 72, 2583-2586 (1991)		
MEDLINE	92013982		
REFERENCE	2 (sites)		
AUTHORS	Jenson, H. B.		
TITLE	GenBank Curator Program		
JOURNAL	Unpublished (1992)		
REFERENCE	3 (bases 1 to 134226)		
AUTHORS	Davidson, A. J.		
TITLE	Channel catfish virus: A new type of herpesvirus		
JOURNAL	Virology 186, 9-14 (1992)		
MEDLINE	92087490		
COMMENT	NCBI gi: 331209		
FEATURES	location/Qualifiers		
source	1..134226		
	/organism="Ictalurid herpesvirus 1"		
	/strain="auburn 1"		
	/sequenced_mol="DNA"		
	/tissue_lib="ATCC VR-665"		
repeat_unit	63..273		
polyA_signal	complement(467..472)		
	/note="ORF 1"		
polyA_signal	complement(583..588)		
	/note="ORF 1"		
repeat_unit	609..748		
CD5	complement(742..3330)		
	/note="ORF 1; NCBI gi: 331211"		
	/codon_start=1		
	/number=1		
	/translation="MDGLKEITAAVASLGGTTDLSTYMNFDLGDMDQSGAGVDDV HPPEVGVDSAMDVAESEDGEGTGRDEPESEADEDPDEYDAIPAAPVST PLSTTRGFGADGLGKGLTPADLAGNGLYAATLF SKTLP LAAVVEOKGALDPG SSVAKSVTLTSLSELNGLIFSR LPTQONKMAVMI LMKVLGRTALP LPTKTDGEV LSA NLYFSGAKRVLVYTHSDITK F I FENDGTFPLTGVDVAP LMKYDPRKLKT LVNLKUNGNGTVEWRKSKGTLWMREIR PVCEGFLRAFPVPTSSDNLYATLDVFK LLKGVSTQSSVYVYVYKPSIGDGLNTKNTSLTLGDAP LTVKLSFRKP NATNSQSGFK NMLTKDVRMPVPSI NEIPTDFVHGNITLP IWLIFAA SLRLD I VAKEM LSTEGDPR RTILCKNDVMPVPSI NEIPTDFVHGNITLP IWLIFAA SLRLD I VAKEM LSTEGDPR SIDVAKLTHQYLFIRLLSDAGIAKKGSLKLTATNNMLI TEMSF IVMYKGSYCWPT VLST I SDNNVNP I P F SGEIPL DMSTSGI VEKVKIVFHTNLSVSRMLLSIDVETAK VHAJKAQVOKQA KSWESAGDAADKAPKRRKEANGTTPKKKQKEEKSEXPXKTK KAEKSTCTAKYTKKEKTKPTQKTSNKNVKS VSAEAGTAVTP DEANGKKSQVQK KKEAAGHTEKKTKKATTAARKVTVDPVSGHRPSSKREYQSEFVSDDSDSEEVISKRP VTTFNNKSTTKKATLSSNDDDDGDDGFTCAGDFTGGSGFLSTAGRNKRSKQK		

```

CDS
3502..4725
/notes="ORF 2; NCBI gi: 331212"
/codon_start=1
/number=2
/translation="MTGRHSPIHQTSPPRPVEKMPRFQREHGASHRYTPMGKRPAP
ANTGSHRGALLGRNGDPERKGEFTAVSKLLKSGYIPARTALLIENVEKEGARRGE
EVMGCSLILPHGDESGRADIDFRDMRSVRGEVASPIDELLFRVGNLLEIVETTTG
RPVPAGLGAVVQVSTVTGIPKMIHVCLDLRSLGKGSFAGKMKKTIIEELLGTGVG
TTNPHCFRTTHSDAQASDFGLNMIIEFLFGLGRVTRGVSPGEGERRVTHTRCLGDDS
FNALFATVHDPTVSRFAEARVTRVQTSYPAWTFGDMWTPAPNNVGVTTPEGSWC
YDIRDLNDPNKMSVTPAAAPPVGPKEPHEEELADPWKPSSTTIGSPAPNLCALY
TSGDSVELHAPLLPL"
4886..5794
/notes="ORF 3; NCBI gi: 331213"
/codon_start=1
/number=3
/translation="MAFSTSDKRIAEILGMSPATLSVRVGVGLPHDPIPWDSDEGG
DDTMTGSGSERALGVETVARPHVYIGPVGICAVDDTSDSDSPSPVTTGAERAAAG
LPRQAPLIVRVSGNERVVSERIEEIGRELARICELOQWDEVDVAGAPVCTVTGNS
PWSVPGSLPPKSGSTTWNPPPPSPSDDTTCEIPSLPSSLVSPNPAPWISPVSD
APLNTSPNMIITPTPGLRLDTPPTPLDADLGDWANDWLYAPQLGQQLRIESGSS
PKGRPLDDPDEVPSKKGPSRRALLN"
5668..5956
repeat_unit
repeat_unit
polyA_signal
6084..6241
6118..6123
/notes="ORF's 2 and 3"
6158..6163
/notes="ORF's 2 and 3"
6194..6199
/notes="ORF's 2 and 3"
6234..6239
/notes="ORF's 2 and 3"
6267..6272
/notes="ORF's 2 and 3"
6513..7286
7960..8499
/notes="ORF 4; NCBI gi: 331214"
/codon_start=1
/number=4
/translation="MASINDGETIGIQVEVPKFRFVCHGCFPPSEVTARGAGSVSL
RVVWVYVSPERDEPSPTRGVDEVDGACSEPTTPEPRFRAIEEMGKIVLVSVCP
LRPALQHRVWVGNAPRRHRSDSVARARFEPWGRASRPTVLTPASGSDDDVDTRTFCG
CGGPGHPVVDVCMCDRQDL"
8557..8562
/notes="ORF 4"
8642..8647
/notes="ORF 4"
8785..9471
/notes="ORF 5; NCBI gi: 331215"
/codon_start=1
/number=5
/product="thymidine kinase"
/translation="MALREGPPTARELPSELVFCVEGNIGCGKSTLWKALMERVAGSG
VNVVEEPDQVNHNGKNYLEISYDPTGCVAPFQNLVDSVNVVQRLQNDIMERSP
MSATVFCVAGSGSGVTPATLPQMAARGEAMRTIATRPVLYLELPPEECIRMR
RDRYTGAGYGDYLRILHERYDAWLSAEDVERVDASRSREETVDVRIELCRHRPL
RAPLTRKSOL"
9535..9951
/notes="ORF 6; potential membrane protein; putative; NCBI
gi: 331216"
/codon_start=1
/number=6
/product="membrane protein"

```

```

/translation="MNSLTIIFLLSGLTAYHAYIADGTCSSSVTAGDSGVVLMVIG
ALLTLMTIPIIGLFGIVYRTRASIEEMRGILQIHLRLITCDQSRNGDVELGAGS
LLTISOPPSSYAEALLMEPVEPQQEGVPLEAERV"
10043..10048
/notes="ORF's 5 and 6"
10248..10267
10298..10326
10420..10614
10863..11285
/notes="ORF 7; potential membrane protein; putative; NCBI
gi: 331217"
/codon_start=1
/number=7
/product="membrane protein"
/translation="MAAVILERAEEVAPGEARVGVPIIAEYVYRALTSDHEMRAFYET
CAVSFALEMLIIVLHASRHPEGSTTRCTDAHTQTEGSTTRCTDAHTQTECSRQGS
MTEFADDLTRPPLGCHGRQIPVLRRLVLDRLRIDYSL"
11059..11152
11499..11504
/notes="ORF 7"
11658..11751
11767..11898
12462..13160
/notes="ORF 8; potential membrane protein; putative; NCBI
gi: 331218"
/codon_start=1
/number=8
/product="membrane protein"
/translation="MKCPARVDIFKREIKTRELEPTKKNVYHGLTGPISLPPNPVS
IRQENWNRLSAVIGFFLLVSTSGSAASIPSPATPDATRESTPGEHRDRALST
ETPTPEPDRGSGTPEVLHVVTGVPVRDRDPILERLAEILAEIHLQLLTPGTGR
EDEDEVFARALAAAEIATIGSVADRVMMKATILSCMLVTSILVFAGVALWIVVARHGFR
VIPHERSDHSPGL"
13197..13673
/notes="ORF 9; potential Zn-binding protein; related
sequence to ORF 10, ORF 11; putative; NCBI gi: 331219"
/codon_start=1
/number=9
/product="Zn-binding protein"
/translation="MATRPVRRPRDRLVYPPKVDRIKSVIFACGHRKRCGLGTRPIG
DSNGCPGCGARDVFAPDLGSGPLPPCPGCGAVVGPVTREACGHVTCNACETEACA
VDRLCTGGGRRLVAICDPYPPCPWRGPRTRPEAEAVQSRGSGSEDACTCAP"
13714..13733
13776..13781
/notes="ORF's 8 and 9"
14076..14531
/notes="ORF 10; potential membrane protein; related
sequence to ORF 9, ORF 11; putative; NCBI gi: 331220"
/codon_start=1
/number=10
/product="membrane protein"
/translation="MTWKGCLLAFQVTLIVIVGVVAVGVALSKGRAPGDDLVLTAL
NELAGDEVDVGRISDAGANITGLRLGLLDHLSRNSGIVAADALYERVLRLRDI
FVNGLVAAFGFWLSFIIMYTCASTRSGKKVAPAPIHSPPTMMSPYI"
14539..15384
/notes="ORF 11; potential Zn-binding protein; related
sequence to ORF 9, ORF 10; putative; NCBI gi: 331221"
/codon_start=1
/number=11
/product="Zn-binding protein"
/translation="MRCIRAAPADGPTYPIIMSPVFCFQCGEAVFPNSRASCXHV
FCFFCTPCECGCGGKLIINEYIALYATKAPFPAPMCRMTAGFWLMPNGCHWVE
PRVVLEDLTAIVLITVAGLVETRGAPEGAVRVARGKEGATADPLLDLDLVEJGG
AIEAGGKAVGVGFIVRTLVELVRYVHDTLAKTFFVMAPEFGSIGALKELLRSRIPG

```

CDS
 15391..16290
 /note="ORF 12; potential Zn-binding protein; putative;
 NCBI gi: 331222"
 /codon_start=1
 /number=12
 /product="Zn-binding protein"
 /translation="MTSPREITVRLSCGHPRLVRPDGEFCFGDGAYCCDCDERAVFFA
 PDHLTVTLKVAPEGLDAUTEKSNKLAKAIVTCGTCGDVLYQPVHERCGHWCRCFC
 ROKCAEGCLARTAVRPLDLERLQESLFGHDTRGLEYSDKFTAEIAGPQGVKEVILLY
 DGAIGNVDVLTGISAADRAEKPIDLVDIVLNPDMSPVCTIPYRLFRLIKANA
 EMGVHDSLDVLTGARNILVEYGAVPGIDRSRLLEFLRYLIVPGDGAETFFALTEFF
 LSIAETWVGDSRCGRDACS"
 16329..16334
 /note="ORF's 10, 11 and 12"
 16354..16602
 /note="ORF 13; NCBI gi: 331223"
 /codon_start=1
 /number=13
 /translation="MAFLPFLFCONQWMSLILCGGCDLISCCRGGGQWQAPRQPIY
 PYSGPMCAHVYPPVPAQPVVRGPRVVRPQGERPVDGLR"
 16665..16716
 /note="ORF 13"
 16740..16745
 /note="ORF 13"
 16744..16749
 /note="ORF 13"
 17047..17356
 complement (17076..17081)
 /note="ORF 14"
 complement (17128..17133)
 /note="ORF 14"
 complement (17179..17184)
 /note="ORF 14"
 complement (17231..17236)
 /note="ORF 14"
 complement (17298..17303)
 /note="ORF 14"
 complement (17346..18521)
 /note="ORF 14; related sequence to ORF 15, ORF 16; NCBI
 gi: 331224"
 /codon_start=1
 /number=14
 /product="protein kinase"
 /translation="MAEKLIIPSCROMFLPARVLREGFLRARPVAVDGGFKLTIGPLT
 VATRWDMVYVYHNEATDNLKGNMPPFEEDCDVLDLFGVACVNEATFLIKKKYG
 SYLHFLPRLHTRGNIEFFSPLSAQMAVLKIRTFHGQAPDRPGVPMVEYAWETA
 HFLDAAPKLEIMESGTRNENKSAQWTFSEGLYGSMEGYFRERATVDICATILAD
 LTGKLALIRKGIYHGDIASENIIMSRSCPCKLIDFESHGPGETHSWYPDRTEF
 WNP IGTCAVAPERSGRGNAGVGIVFQIGLIALINWERMVFNHNTWKGQY
 RAHVIVKIRACTIDLRGGARTLARVDELIGIVARCLERDPAMRP SLETLVDFESKI"
 18557..115670
 /note="unique region"
 complement (18768..18773)
 /note="ORF 15"
 complement (18773..18778)
 /note="ORF 15"
 complement (18837..18842)
 /note="ORF 15"
 complement (18854..18859)
 /note="ORF 15"
 complement (18866..20008)
 /note="ORF 15; related sequence to ORF 14, ORF 16; NCBI
 gi: 331225"

/codon_start=1
 /number=15
 /product="protein kinase"
 /translation="MAAVNWLKDEPYPEKPTRRHLSFGPARLPTGDMWIMTYKPE
 ARCLMRLTNPNLWSGPEDEVLLGLPAGVPVTEKIFVKEVYPLKGFQMEIPKVGAGCL
 LFGFSPLSRTGMATLKATPVRFGYAPAGPIPMOVYAWEMACKIDAAPRLFKWELIV
 GEYFVRLATISECSAGDVGYIRGGRPISIEAAAVKTRTELASTLYLLAQNRYHGCV
 KIANVTYTPHGRGLIDFEMAHPLDMTMSGLRGLEYV IKDMVCTDEYRAPEGHGP
 FPDIISAAQLWQVGLMVDVIGIEIENRNOGLWEQDPFGIARAAREVIGIRSLILAC
 EHRAFDLYITARGCLRTNPERPRITLILIAQLTKFIREVATQPEH"
 complement (20123..20128)
 /note="ORF 16"
 20139..20161
 complement (20171..21334)
 /note="ORF 16; related sequence to ORF 14, ORF 15; NCBI
 gi: 331226"
 /codon_start=1
 /number=16
 /product="protein kinase"
 /translation="MTSSCFIRAHPGVYRRCGFTRGCGIKVPCALITDFGLAWVHAYY
 TDEALNWLADNPLMKSGADVMDLIPGDMVYSEHVKKILSRVGFPEFDSFIPDIRV
 EECILFGFSPLSSTGMALKADILVGYAPAPAGPIPMVYAWEMAYKI GAAPRLLRW
 VLIEGEGLEQFATIAESGLRGLRDYICTHESGCKTPMDLGLVVKVHGLVLAALGLML
 DSNLYHGDILKDLNLTVPFGPYKLIIDFPAHFPEARMHPIEARGPITWLPMPGPAC
 NPPEHDEETPCRRRMKEGLVQGLIMLETITVLDEALVRDENMEKRPDKRLVSA
 VMEENGSLAPHDPRILLEGYLDLIGECILKRDTAVRMDLVITIMTELSLLIEKFKL"
 complement (22737..22742)
 /note="ORF 17"
 complement (22832..23728)
 /note="ORF 17; NCBI gi: 331227"
 /codon_start=1
 /number=17
 /translation="MIPPPIPVQLSFPICSRISISVPTGTHNALDRILSRPSPKWK
 FTDYINGALACVGAAYAKALDNFGMRVSIVIGGVTVGLCAAFGPKKITELNALSQ
 MLAEATEDIKRERLTKALTSVITLAASRAPINTWTRPRERFDALTGIPARMDDALCE
 MEFDTCGSGERGLHALHLSGVIVGELERLLAVEGHVSDLAQVYLTMLIVTLFERDDE
 FGKALLKVDSTIIALLESRLDITVLEEIQRVMSASVTRFHGVRPGFTEAFAMHDAIN
 RCGLYGTARGEPEVEEFGSTV"
 23762..24718
 /note="ORF 18; NCBI gi: 331228"
 /codon_start=1
 /number=18
 /translation="MKASQERSAARTASHVKEKKYMWASPRKRRDLAEKAAALWL
 EPSARICLNPPEMVTILLHYAAEYSTSRQYRGDFKGEPAFRFSPHLDPTVNRFTIP
 PVSVLNPLAVSHTLLHHGGLAMCPRVAFTENRSKRPAQAYELTVTFNEARVP IP
 QRTEWELATLSPEKTTVRIINWSGTTEDQLEITERVLRSWADFSEYQLSRHSEF
 VTLDLSRGYDSSGQYHAKMDSYDEVGVYVVRNLTRASIPRRFTLVICHGTAVT
 AVPVELTDTFYKWLADDGDDGDDDDGDDDDGDDGDDDDDE"
 24638..24712
 24748..24753
 /note="ORF 18"
 24842..25456
 /note="ORF 19; potential membrane protein; putative; NCBI
 gi: 331229"
 /codon_start=1
 /number=19
 /product="membrane protein"
 /translation="METMCPMKSIWFWVILVSGEGTDAVEIYMLEIHESAVPCVG
 SKNWTMGCDLLAGGILGCGEHLHRDPTLRGGVINKYSYRANPRFWPLSVMTILH
 VLIITACTLIYITAKYRSRCRRPIERTPKDPPDERVRLYLDVRRRRYWP CSCGEDD
 RTRLMPGDRVEYDGREKLVRSQWVTHLLTDPGDTYTIANE"
 25452..25457
 /note="ORF 19"
 complement (25562..25567)

```

CDS
/notes="ORF 20"
complement (25576..25884)
/notes="ORF 20; NCBI gi: 331230"
/codon_start=1
/number=20
/translation="MASSDRFPJFCCYRNCIFRVSVDVQVPIAHLRDHFGEKPGCTSI
RFRCPCEIILRRPFLGETVPDENNIRRAHVALIQNECRSPETQIMCIHFLLDPDES
"
CDS
26089..26979
/notes="ORF 21; NCBI gi: 331231"
/notes="ORF 21"
/codon_start=1
/number=21
/translation="MSATGQPLTEAANGCDDTGLCKTKNHDGAVPRTTCEPRSLICE
NRASSVLTGLGPRGPVWDPRIREFHGVESLPLTDTWPGPPSPMDALGHGWTPS
GLKWCPPPPPTESRFSLELLAMHQKAQRDDTVTGETETAREDDVIFVEETNYN
NPDIETVELKDVTECHGTIKRRYPMRVRRAPKRLVLDQVVDYFADSDDDTDAES
DDMSVLSDTCTETDDASVSSCGSFITDGCSESEDSASDETDDSDFDDELTSSES
EESESESESESESESE"
26926..26977
27099..27104
/notes="ORF 21"
complement (27145..27150)
/notes="ORF's 22 and 23"
complement (27146..31357)
/notes="ORF 22; NCBI gi: 331232"
/codon_start=1
/number=22
/translation="MCP IRKPVNQWELSENVRNLCGTFGEMVQVSTSPQAAAAVF
RMFGYLRTLFOHLEITLQSDGAGDPVRMDVLIQIEILGTTVSDSDSVFVEIAHLS
KYKIDFNGCSVTWIAOVKAADRAVTAAGVTGGMGAMGTLDRAMIDIKLDLV
LPTPTGSPFVGVLPALPLPVPYTAGSAGSPDLDRFGALFLGLTPTATKK
VFLQNDILNKVAQISVSTAMNSFOAGLAKYKPPRPTADGALTARGIAPDRPQ
AKYETKVTATNMGFDVLIGKVPCKSGDIESLVSGILLACKSIDELGALVGGISDG
AIVSRQMDQVRSILDSLVKTIYDTRDRITAMTGCVHIVTEYANDLLRLVLDLIAKYG
EFORTISGLEPFWATIVSNTEGLPLVYAAAVAAERAVIIDEFKQMDHANNKVEV
EHIERRAAATOKOTEERRAVDLOKAVETTAKMTDMNGVEKTIADLRQSLVANNLSI
ESLKGTLKGTGTGTTVMOLSKANRQLESRAELQGGKLGIQENEAKLRDEQHKQAQO
IANLRDLPATEARLHKQKTKLTETINTGVSTARVELEORVELVAEKKSIRHENR
MARUKDPEAAKGPROSERPSOTAVHACALIKNLRDITNKRLLTNTNREKSI
SKKELISLITKGNALSQADAREQHAVIKREYDALRERGGTARTKKDALKALAS
KLITQESQALAEKLVMELEAAKRLCKSVTSKVEVEVRVNLITTEILARTKKTVD
ELKLNNAVLTAKORLOARVEALTDMNOSMKLLVNGGTTIDEANKRIITLTORLATVS
GELVTATETLAPVRGSKRVNRHQRDELAKESATKSCENQTRALTRHATEMRTAAG
ARDGLEALGRKRLTYDTETKLAGVERDVRVIRDELTRMOKASHIADRLQTLHAEKV
AHDGMVNLAGLTARTLEIQRELHVHEKSRDAEQRHVASEAALREQIRTLTDERO
RLIATGVADLORENAELHAAITEAANTCADTKIQLENIADYNGLIAEYERLGRPP
GPVDPATTEEPJAGYGGGMDTSGAARAPIDHDMFGISLLGESQGGGLSDYHGY
AGGRLSESDESDVDVPDNGTNGDGVGREDEEDRGGETGESGEESESEES
EESDEDP SAGCPLELIGVPAGGADPEPARVPSPRPVTLIERGAHVAPKPTR
VRGRPAARSHKVRHLSITREYLRALAAAGEPWSAEEIFQRINTDMETSSLSLNTV
TAADIPRGQLAVLRDPAVEYATPTGTPVSDRDRHQRVLSMKAILLIMICAKHGIMNHE
EFTDGDLSWFKSYDNLSVTGSDSE"
complement (31281..32537)
/notes="ORF 23; NCBI gi: 331233"
/codon_start=1
/number=23
/translation="MYIRINESRGESTLNRILSASRVPSSVSSARFLSTCAESSRL
RVSTSLRSSRSLSFASVNSREITAAVPVVTIAGMEDVITIGLDDTVYFLWALILSN
DVLISLRKGKVTVDVTSNTSVTIAPVATMDPMERLFWTEGHGVFMDLKGAT
YEMLCVLA5TAGVLSMDPMP SLLDRFTDSDELEIFVHLGNTEINGATALWDRTVR
YAFVPSIELRFXQVNNAILKGGDPBSYMDRDAVTVCSTVGADALVLPDGGDGV
PVAVNRADRYLEELSERSTNRTVELLAAPTPPFPINFGICVLTRVKGGLCEVFII
IKDSRNMWEDRYTFAINTSMCIGNVISNGLIYETITLAGIERTLIQPPARTSRPFCVV
RDGSHTEAQPVGALCKQAKPRNL"

```

```

CDS
32293..33294
/notes="ORF 24; NCBI gi: 331234"
/codon_start=1
/number=24
/translation="MTSSIPAIYVYNGTAAISRELTIAKRERDELAREVETINLELSA
QVKNRALEETLGTDRDAEISRLKLVSDPSRLIRMYITKNOTLAQVRKTLLEHLARE
TKKSGSPGVAATGTGYGPGYPRVP1PVADYPIHYSRVGPEYVLTWYAAKCKQKN
LIDTVRGVGVTFEYTVILFSASAIFFVGMFDRQFVSVTEGHDLDYFCMKVALIRVTTQ
YLTLINTSDSAQALAKTHNQQLFQTTQFLRVQQQFQVLDVNIIRSAELGELVTSAP
PLDEYTKRANLITAEAGDEDLKLILADDFIIRVPAEKNKRRRAIGGSEELTACVE"
33342..34838
/notes="ORF 25; NCBI gi: 331235"
/codon_start=1
/number=25
/product="SF1 helicase"
/translation="MDVEIKEYAGFQYONVWRGVASLMMTHPQGAIVLSGDAGCGKS
HAYQQLIEASRMGFAECVNITATTHKAVGVILQKQPCSTTYQSSIMASIDMQANFEF
KCLFDARHQAIEWMDRIVSNRTGLAAVRKKGHCVCVLIANNRCPVCVATIHESIRCTDQ
YPYFIARTWFIIDEYGLISELSLKLILVALSKFKLPQCGYILVFTGVSQLPSPHPQ
IMQPTLLRDMILINTYSLFINERVEDAAVAEAINLFQFNVITRPAVEMINARVIGANAV
DAEYCDSVTRIFNGNHERDAYNTRHTWMEGRGAGRVDLTPVIEHVGGSSACREGALI
EYMCSTVPKIFNKATKGRVTYGGGVFLPTQERYTFFGLDDVNTAVVGGEDRLYL
ERQITKYKQWTVSFFIIPPVAAINTYSAQGETLDRIYVAPPEKNYSMSSIRASAYAC
TRVTRTAIELSCNSFSAKSVGRTNMPPEKLINHKRSVEMGVTPRECPY"
34727..35146
/notes="ORF 26; NCBI gi: 331236"
/codon_start=1
/number=26
/translation="MOLVREKRGAGHOHVPKRTTEPKQVRGDIQPERMSLLMMINPFF
VDSVILETVDALRGYVQQPPFPDVRNDSRKYIMDRHPIFWFTNCERPNTLTYL
VSKIRIEDAIVFGASIVTTLTHLVPQREVRNCRKIKII"
35148..36014
/notes="ORF 27; NCBI gi: 331237"
/codon_start=1
/number=27
/translation="MA5ITGQLSPMYINNVTSAVQYIPVLERFDRDLEVPKGVF
VPLSGMFTFPVHGM5LQRRPLGIVQTGGLMSEPMTVDVVP5GKLYTVLNTGCDTLRV
GDHFWPQPPQEDTDAQKASKPGTRRICNNAASFNRMILPWGFMVAVKKIDP1VPRE
LMHQLTRDVTABSSSELLYGAGVYDLEKTSAAAGVFKDIAEKKAVTKEEVAANT
ADQVDIFRKIVDEFSTHLCNVNFATIGMTHSSTLANPTPSLSQNVQWCVNCTGVTPEGL
QFHAFIKQFSGV"
35991..35996
/notes="ORF's 24, 25, 26 and 27"
36035..36040
/notes="ORF's 24, 25, 26 and 27"
complement (36075..36080)
/notes="ORF's 28 and 29"
complement (36102..37874)
/notes="ORF 28; NCBI gi: 331238"
/codon_start=1
/number=28
/translation="MSLPGRGTVK1ETREIRIIVRVNGETGVYDTRAGSFETVSCQE
FEAAATVSVVPVCDRFCGTSLYEVP1LTFGTFVVGTCCTIFSPGDVDDRSIPAHMR
KYQOPTEAQTVMQVAPGTGLKYSHIPKGVGLGVHVMHEDRVYINNIIGTDESDTG
KRCVETFEADIPCALISGLTLDVGLNEVILECSVWV1PARRGCHAKLETRDTVSDGLE
KFCQSHATLPTLLASFSTSSPERKTFYEAHVDA1NNY1KLLTIY1SHKGTETIE
QYILEGSKLYSELI1GEP5RVIDATKAAQ1AEPQTHTGADRQRPQRPDGP1YSPDR
PMFTGYPAPQFCGDPGLVSHYNPFVPPQSGCGYGPFRVGGVYPPQPNPYVLP1SYG
QQY1PQHPQPHGHQQRSGGDLKAE1IET1GL1APKTNVQESLKSF1SE1LESELKQ
CGIKRAAGN1ERNCDYDEEPRTRKARPTEKTA1VRAP1YGFDS1TAL1TK1QV5
DTVEK1NKV1IETLL1TQSSAQAP1LSTP1QAAP1QPS1LQPVPEP1APQEPPEP1PCT5AP
TLEAS1PQKRPV1VSKGAFET1LNI1"
complement (37738..38499)
/notes="ORF 29; NCBI gi: 331239"

```

```

/number=29
/translation="MTRKIDAPWLIQWRAEVLIPVWGHSHAVIEDVNALAHRIENIHH
RSQSRMQLNGDIFDGVQLTILLDGLGVCKLISLVPVGVRLGLLHRRKKGICGCTRK
PDRLGLETYFGLPEEDHRSRHFQFNDFGYLLRSMDLDAYTAPITAPVPGVDCV
VIYKVVHPIPIPAKPLILALAVOVEMTVTQQLRIRTPKIRNSVVSROVTTTRKRYR
EDRUSGEDIGAESKRGNGSVRYTGRELDSINSRI"
38368..38994
/notes="ORF 30; NCBI gi: 331240"
/number=30
/translation="MMDIFNSVESINILNNSVAPMDHREHEGAPSDKPGWVNSFH
DFLWNSRSDIITSVQVDRGSCPCASKNDOCTSLAEMESWATRLQDLSAMSL
SWVRPLANWVINYHFSHGICODPIEKLEYQIDTWIDVHYRAGMALGRDFAMWYANG
SQVSPNEKNICLSSLTTLGTLKLIQVRGDIRKMTTSKKNRGCTAAPL"
39010..39015
/notes="ORF 30"
/translation="MYLQWFFPCPSFSFHGSEVWVWVVDGRTGCRTHRGRCITV
IQSAGNPVHATDYEAWLRMLGMRIGRLVILYTSK"
39359..39763
/notes="ORF 32; NCBI gi: 331242"
/number=32
/translation="MGRATTDSDTSVSRALPAPVPGSLFLPWLHVSSLLSDQLM
GSKSDRPLRIGIGVSTVATVPTDTPMAHTEPVFLGTRSDLMFSPHNLKNSVLQ
PGNLSLSFLSSFSVPRNPDPNFASIAVS"
39821..40693
/notes="ORF 33; NCBI gi: 331243"
/number=33
/translation="MEPISYVIYDYTAFAWMLSTTTVDVREWDRIGGFNLTASN
GRLLDLSRRAYVHNLVHGHFSELTGVGGVETPLFRALTRAWNSEPERVLTIR
ELRTEKLHGRGDYGRNLEYMCTNLAETPEIFETLCTRILPELKLIVTTMKVYIT
AMYAGRDDVLQVTFYIRRSLEVDRELLEFIEGELKHSCLIPISRKMTDKTFHVSGF
ITTTCTKELDRDDTARLVRHNGTGFETVDRIVMGATMMVRYRTYRAVESYDEFWYKY
CGWMDPTGELPDE"
40536..41861
/notes="ORF 34; NCBI gi: 331244"
/number=34
/translation="MERVLRLQISTVSGGRCGGIEERIERMSRTTNSGNTAGWIP
PPASFLMSTVLRVHSGSDGFFIHWKHVNTVNRQHVTRVMEHKTITQSNVVKRAT
AKFVTDKRI SNVGSKSTRIEFTREHSANYTANCKPLVQPTYKSYFNLLWHSQGE
CTITHRVNNIATYTLGVGNPMSVWDSMKQKFLTPFPFSSINLAGIETNQ
LEFISMTINTKLTHTYTLQGLKTLPTYSLLVEDTKENFMFNIALFLECFINTPNKI
GMIWMNERYLENTEATSTETWNEYKYMLEKCAPLNGFSNAFAQRTGQFYKN
CEWVHAPTFVAALAEAVLSYGSFLINTRKIKSFKELVWMLSVTPTDPLRVLVTQDD
RTDYRTGDRVILWMSHADFKVGSERLVLGKKINQYIKCKTDGHDGQD"
41707..42423
/notes="ORF 35; NCBI gi: 331245"
/number=35
/translation="MTAPMSIEPTGSGFLTELISKWGRISDSFLERRLTHISAQQ
EITAIKTENEARAKALROSLEDRAFWTKRITALGYKVAKRMAWVTEPGESTAPHL
EWHSHKGLLGLFEISAGOLLCPSAITPFIIDRGPGCTPLCLRTFDOKLEGHLYTE
LISSGHVSRSRGDGNRAHFTGTPVILPRKTLGKERATLSAAIKQFKCLQVDHCT
LEGITESTAIKRAKILDLICE"
42425..42868

```

CDS

polyA_signal

polyA_signal

CDS

CDS

CDS

CDS

CDS

CDS

```

/notes="ORF 36; NCBI gi: 331246"
/number=36
/translation="MDGIFPSTDKMLYENVTIHRASGEYQTATLISHNHALPDSGS
PCDFRFRASIKKNGISLTLTRGPVTCIKFQIPEDKVISLETIACLETGTTKVVGDYCI
TVTTGDVLEPPTPQAPPTPEAVKQPIATRKRSGLSALLMPF"
42950..42930
/notes="ORF 37"
/number=37
/translation="MSGVKLGFSALISYTCVFNLRKKPIMMYRLGHQOQHVPT
DISIFRQKVPYPTDAAIANKLIMITRETYGIPKTYIRLISGTFWAVIKRNIYDNDR
VFGNTTSEMNGLDLVYADDCTKQLLEMPPEHVTDQHLGAGMSKSLGRTGFRHL
TETFRKMLMSVLTFNSPVYPAADARADEACDLVTRATVDRLRKTEETVTLADREF
RGLLGPVWAKRLUKVIGPIPDPTARSIGLVFNDIRPKUSTEAARTLGGMLIETLRQY
SGYLPAVALVDLTETVWVHVHVDTCILMDKRGRVFRALGEPISRSQPLSSPVARTSE
STMKFQALDDYMRASNNAKRTVTAPQTVAAQESRNPFTRELDSCQLSRATENLI
QLTGQNLHQRTAIAESSGLVAMPWTSPVNTALAHMLATNGAAIQESMLDDINSRNI
DLTGQRRFTDJAKRVGLLQSRDFETLAVHVGELHPDDRELVLHALGPTSTRSAAQIDP
PEIHLRPNKTLKLSFTTSINLPEIEELHLWTITWETVFCGRNLFTRRKHALQYTTT
GRHDNPEQSIOQMFVDSLNPAGVILMDRLYSPKTYKLITFTELAVLSEKVEYVTTPEIF
AAQICLNPSEVRLKNDACEPREIYTVGNHELGVGDGRFVAMEDV"
44891..45256
/notes="ORF 38; NCBI gi: 331248"
/number=38
/translation="MSVLCTCCCKCPNLYIAMGFFRRFTKTQVVDIILSANDPLTP
DISMRADVHYQAARLTSCIQHPVAVMGTRDICALKRAATNIAAGNALMDQLKRWTH
DRAENSTNLVGIINDINS"
45306..48677
/notes="ORF 39; NCBI gi: 331249"
/number=39
/translation="MEATARLDNALAKIGSGGTFPDSGLSHSGVKWFIDNALRLBQF
TGSNVNMEYTGSLGPNRDFGQKYVIVPLIKTISIGRYLNRAQAGNLAAPFLISSTF
KPASLPQATFTHMLFDITEORVAOTGNLQTVGGGTTTTYDIVMSLERKIVTTEIDS
LELFYNDFLVIAESITRELESQTLALDSVILEHAKFCARQPLLEGMCGLREARQN
DLTDGAIIRACQLENIVGSGINSRPNLSLLHTALKLIDTSPKDAAVVTVTVFQNS
SLDRRLVESQVGGSEVYIYSNGAVGMQPIAKFKTVPPEGELRAFNPVPEVSPGV
ARSAYGGLKSDVDVMAIHLGNNETIPVPLDSTKESYVGNSTVETAMTQECIKRLETT
VGTCPNIYDLSLEPGSFFGSSGTGEEITPAMVNRSPRSTIYMDLDRGDRVOEITLRQ
LHARVLDGKDAFDTSFARKIARLEALEKFNAAEVANLLEPEMERKADPGSIFE
FDRQMLLEIPDAEADVLVGVDTMDRABAYAVPRTCMYNSSLSMKENEMIGVLS
MIQVLQSSAATNYCOLVOTALSDESAGCAVNNIRHMDLTGATPTPLRVWOKITL
RGRYVPLSDPYQCLAIKTFDDPAITDVGVRVTVVWTVWNTIRAGFDPAIVNFKTI
FDEPDSFTFSVNNMKSTCVTQALATATELNFMTAVAAILPITDTLASMGGMVDT
ETNFMWFPVYRNACFTGSGATLNPQKLMWELTGVTKPLLVRAVWKCTRITMKFTG
QGNPFNGLYHLJHLSDTKTVILKHFDVYKYSYGLCVTRTFTGHGSLVJPKSA
RFTITGNMAYGMPISGCTKHAWTQCMELAVGPAGLDRLGLYIPNICTDIHGMGVFN
PRMAYALIFDANGFTPETTKTASPIYINILASGRYVEGIGAPPDVTVGLSOCPTL
VAGLGAFDRLSLVRHLNLPDRAHYMLAAVDGRNVFETIPVIFSNVYMRDLELL
KRTVDAITADRNLYRETMLGIFADNYVIGKNGEHRANTKTVPPTTAAALSPKVLGES
LYNGAIGDLEYGEFSKLAKYTGVTTSIRIDINAV"
48745..48750
/notes="ORF" s 38 and 39"
complement(48788..48793)
/notes="ORF" s 40 and 41"
complement(48880..49290)

```

polyA_signal

polyA_signal

polyA_signal

CDS

CDS

CDS

polyA_signal

polyA_signal

CDS

```

/notes="ORF 40; NCBI gi: 331250"
/codon_start=1
/number=40
/translation="MTFPPSSLSIAPLSTVFLFTDNVRSNMFVCKIVYRCLSPRR
IISPSICAFSWANHGRLASLTAMRRSSSTFLFMASTRMSSSTVCFIFVFCVSR
TAMVWFSFSESTSLMCIISSSMRLCFSITSGST"
complement (49423..50325)
CDS
/notes="ORF 41; NCBI gi: 331251"
/codon_start=1
/number=41
/translation="MYLRDFHELAPSOTGKGVEYFVVGFTKAGLKLGRLLVUNKRYL
IQKGGVTPINKWIMYNDQAFVVDHREYLIQIFRSIPFVYNQICLTHRFQVSGSF
TRELCGFATIRNLVYSLDPSRLARLQGLRLSLVDHLKTTFPDDYLDLRLPSPG
GLYTTSTAEMNPTGTFDLAAFDALFEGGCGTHKPPIQFTYVRFDKCEGTCENADH
IHELTRIIGEVRRYIISNAKQVIPLLVVLAVITAENRGVPERVARGP IHATGN
IHFDTLSVYVQWMLKGRVCAP"
50324..50713
/notes="ORF 42; NCBI gi: 331252"
/codon_start=1
/number=42
/translation="MASFGERMSRAYTMSMGECYKASTSSSQRRRGVRDDTGTOP
HGLVRP ITPPEMPTTAHRRPVVVEGDRGPPRPEKEPSTTTTKKGGPPTAA TRTTSKKI
VKNKASMDLACTEKELEDECSPLFLEM"
50747..50752
/notes="ORF 42"
complement (50977..50982)
/notes="ORF 43"
complement (51032..53707)
/notes="ORF 43; NCBI gi: 331253"
/codon_start=1
/number=43
/translation="MDALRNKTI GMDLIGAHDMDRFDSLHETAAPLPIQGPKNMKIC
AFINPETHGLIKVYPEHLGFFLEKIGTEAEIRAWGATATVMDSDVSEVYTPDDSV
VSQLPTELLVEARRLDLVAASVEKRAVKI GGGSTGVSTAVVLARLETRVSWPHNA
FLELAQLYTYRNYFGLGDFLALVATSLHLNLTWPLAVTRKI TTPLFVDYVQEEVE
MEVNGEWLQSVKPYTFKAFKGVLETDSEGLSEDAADDFLAESI TLKLEKNQWT
KMDORDLKLTAI LSLSDENPRIKRAKNNIKLUVDFNGIDRMAQPAFVKLAKLMGS
AHMVAI LQSTLRSSILSGNDVAGNTTMSDPELLANKLRQYTRVLMADLSLKVR
TPMGMSVSDNSVKLDRYMGKKDADQDKILQFYNQOGPTTF IEMLNKIGIESLE
IPEKREP KGVFVNCLSVYNSTETETEGTLVNSTGKGATOSGTKNKSCLSDLGK
FRYVVRATPGTPPASILGLMLHILRTALFGVGVCS IHMA GEFFAEENLDLELNNKAA
TTVMKMGIDQFRCGYFNLGNDNDKVKKSFDMIEFVFCAGKRALLSKNGCNGNLRSS
NAFNGLEWLRKKVKAPELANNMRRARI NLLERYDVQAMLLDQAE TRSLYILRPWMT
PFKTELAQSTII LSKKADFNNAPTLLHVNLTSTVITSPNTHTFTLP LAPIFNRKWDQHYV
GII LEAMTVYVSQATRTGTATGGGQFGADYGLDLSGLSSLSQVSEDEPVTSELINM
FSLDKRYDTVVERKALSRYEETEPVYGEDVLVLSRELEQSDLEDCGTENNEP LIVLQ
PFEKDCSRGRAAATLSELMAGESTKRI SEGGDED"
53657..54664
/notes="ORF 44; NCBI gi: 331254"
/codon_start=1
/number=44
/translation="MGTADEIHPDVLAECHFAHKKINGLIITONEKKRAVFKSLKSE
PSETDLSIELDPYVSFRKYVGLAADAPISWRHRLITLISFARFLAREPEASQYI
TORASTAVDCSGFLETCAPRMKLYLPRWKYIGROGLRRKFTQCKTSLTPSLYNIK
LOVVVDIMRYVTEDP LDLLVTIYPEETLEELAKVKRGAANLEHMLGIYHVLNAHI
NWFARFRHGLNPGITLPSSELSQELGWAYDLRSCGHRLLS IRLBKTNPKRMKI
YYSYDSHGLIISOCKASVUKAPLSIGKLTALVINTFPFKQYHICACGCEAI VTNDDIPVS
E"
54642..55364
/notes="ORF 45; NCBI gi: 331255"
/codon_start=1
/number=45
/translation="MTFQONENWNPILSDISDIIVDRTAAQTPTATYAGLLEADARAI
RPVQOQYVNGPCCSLTWAGVPGVYITQFVPPVTPNRYRVAAPGCAFGCEWCEPPT

```

```

VSAPPPPSQFSDEP TSPELAPAVPKIDHEAPAVATVSSPTSPRITTESRSVPTKEK
WGRKRVHKKTHAEATWIPQIILKRGMSLSLLPAAARGQRSASVRRRVFTTNSKDVLYYD
KYSAPVMIQNTVTVVEKTSFK"
55360..55365
/notes="ORF" s 44 and 45"
complement (55397..55402)
/notes="ORF 46"
complement (55427..59494)
/notes="ORF 46; contains 23 potential N-linked
glycosylation sites; potential glycoprotein; putative;
NCBI gi: 331256"
/codon_start=1
/number=46
/product="glycoprotein"
/translation="MKRTMLAILIPLVYVAGETVADDEVFFLSAPQLDGLTLDMSG
TDAVITGSTLAWRLMKFTVSATHTLREYGRVAKSVNSTAATKCTHLCYPGTVLKLVS
QATRSNANNASMRFAFPVITNHELNETCIPMEYDCEVISEKTSINTLSITTIYGSN
DIKKALGENSTCNKWTQGCALATALTAYENKSVDAKKMLI LGTSTYTCLLGTLGIG
ELEPNPCWTLQGFMCYGP LAEQVCVTAGRNPFQGVPRADRPVSTGDCRVCGRADILQ
FSLGTLKGLFARGCDRVFVCDLIDPYDPATAAEFELHVECSRFSRVSGPGEDEM
GLELAPNFGLEGLIRTKQDLRTITIGTIDIRNNKNALVINHIPTPELLDTFHFWA
ELDFTFRKTHRTVPLGWITGASFVFEQNFKLLSGVDMVYIHPQIQWMTSTPTLTFC
PMTIRSDSLECCGEPHQCPGDFGRHLTVATRFLLGHISPEKLTGLDVSQQLAGRRV
RQDYNWTFQDNLFSEVQCGWVSRYITDVIQPNWTOALQINGCHGVFCPELDLILPS
ADATPYVGLWMSQ SDLEDAVDFSMALGTRFSIEGALGDYMDSSENIKTSTVTHHA
VTVYRMAJANMEELGEGLAADFESMIWANVTRSTGATTCARKRRALTQGGDGVN
RTIPFSASVQSGTSSIVRGTCPRWCPCGIIASVRGLLESRTSYQGYRLTFFQNLVYT
QIQLNLTGATKPEDLPVITQATKPVSEFADINTFEGVNGSEYVYVGLTGLDVLQYS
POVKSCAFINTNETPOTETITIDERFETTPRPVADGFVIPAGINFGSDNIMTVDY
IDENVTCMYNSPVSQCIATVCAANVTECTPRATILCNQATAAILLDFQBSNELLKNS
LVLDLLEHKVKMFAPSGSVPTSDKFGLSVAATITMSAALVASTAALAVATLAASK
IDGLAQDLKDTADVITELGDSIALISAKLDNRINRAVNGRVDLQNOINQLAMLDTNF
KRLATGKELGTTTNERLCEVMAVYQOQIMSLTNQVTOGAIOIGYKVMIRTCYKS
LLAGTAGCPTDASFRDHPGLTFRTKTRALLRDQKLFIVNEVPQTLTARSVOVFIP
DPTITEKVCWPDYKLHMDGVKVPLECTGKCYSEP IEATDYQECITNPSCTRCYCG
DCYRGICVNRITTEDISIKFENVTHALSVDLTSPLENSIPOIVSMEMEIQDILKLELIQ
LQKINTSVHMENTIGDIDAMKATIEYRAEMAKLURVTGEGMILKIFYIYALIGLVIAGI
LIIAIFPMAVKCYQARALLSMTAQVPVTRPMGMNY"
59028..60584
/notes="ORF 47; NCBI gi: 331257"
/codon_start=1
/number=47
/product="serine protease"
/translation="WMNERVNEVFSEMTSSQSYSGIHVSFNSCIFVITGRANLHDA
LLLAEDRVACTSEFNTVPCTHRWVHLVAVELTDFATRPYSLSVCVADTVKVFHNLFI
SVEPVTIASVPQESISVPSNCGALRKNTSSSATVSPATATRGIRMTIASIVFEITVLVD
GVATGSPNALVDTDFCRGMKHGCVQQLTSGSTGANALCSIKGVGDFIVNNFSGNVP
DWLPSTAPSYAVVSESGCYDDGAACDPTTGNFRDHRGRVAREGRDGLGTLIRPAG
NGGPIDDCGADGFTQAIGTVTVTDTYTRSERCAAVLTVPPNETVWYDKCGCFIPS
SSSARAPPILGNMALLIARAHTLTLRMQRIIVRAAKPVYTYGRGCGHMLNRVDRW
THNRFGEVSPSRLEIARRELSTSRAPVAMSTLDTCOLSEVWVRVRLGATAPAVFR
GGYVTEISPSGCTIIEILCKRPLDFSRDEFEGEVTP LFWGEPGRCKWTVSCTGGVIL
TTGCTCHGIR"
complement (60645..60650)
/notes="ORF" s 48, 49, 50 and 51"
60753..60758
/notes="ORF 47"
60771..60776
/notes="ORF 47"
complement (60812..60817)
/notes="ORF" s 48, 49, 50 and 51"
complement (60845..61681)
/notes="ORF 48; NCBI gi: 331258"
/codon_start=1

```


CDs

```

DKKWMDCDLVFGTHAAQAKAFKAGNSVLGSLGNVYTYGDHWSNFDNGTVPVQNLN
MCSHYVYTRKTDVYAPWHEYDYDQRDQGYLMSLPAIIPRCKREGAFDIETIVHENAM
DODLNQCKFEKFRSEESQVLRQREAGVTGLPSPVGIQOKLHEIVSISLVWC
NYHKTGPKKKEYYYNTKMKENPHEVLHLDASRIKEACKNEFYHLLAIINR
LRKSVNVLVYVNAOFDIOVQORLYRYAFKRAPRCKGHDDIPHEKGAIMEKWEAF
LSVPQFLKAOIQLMGQDILKANYLKLLEGISVLAQAKS'PMAKCTTKERIDSYRQWK
DTQVQKSHCFCCDIIDIMVYVCKKEFEAKDGSINTVAQLIIKFKDPKATPIHKMD
DITDKLDGTYRAGGTAKIACGLINLIDSLVIRIARLAKLPMEEYIYRQLACTINDIA
AHTRGVNTCGFIOSTKRVESVRNKARLDAGIVMATDYIRNSLFTPETIPRGGFVMA
PLTGLFAPRTQCELCIDFTSMVPSMCDLMI SPETIVDSKTNRVGDYMGYDMSKI
DOGEFEKTLVAVDRTPENPKLVHRHSDTSLSLKRYLRIRLTERHKRAKQSGSVAEY
HNRQNDMSKICNTHYGVSEHTCSLMIITQOQHKIKLVNEFIKTLNRTGHSIFPNYGD
TDS'TMLYPSDESETQLEDMVTLEDEMRACLREYMLKKLSAELVNRVKEKTKRTDTFV
QSF LSDVETVLFDHWEKLRLEFSQGEVIEFFKDGCGTMMVDDPLTGIIMDCSTPFSSEL
ICKLEYENASSIGCHVAKWWSIGSTYLFKKXISLYHVRVWRMCADTDGSPSHLYFV
SLSRTRAKORGDH"
76809..78608
/notes="ORF 58; NCBI gi: 331268"
/codon_start=1
/number=58

```

polyA_signal

```

/translacion="MFYEGCVPTLMPDPTSIPOYSLSVHLEANGELINTKIKRGM
TGKSSRFATESIINDFDLIFNGGALVKTHLEAKPVYTWAKLSAPADILHSEPP
VYENGICLANWSITLIPHRVATVRLPCPDIKGGELVVDLHESGKSSVLIVLAD
GCALNHELFSANVLHRELSFETIMAFRAFVAGGLDPNSIVFYERIPRUKPEFLSKI
NYHLEFGKLLKIPFTVLHKKRTTIOAIVKEDKQIDYPEEMIKQDPANAMFRRYPID
LRLERIMDYFGSGLKCTLATPTPKYTVSGERSHSIIVIKNHDVHYLINDSKFLAH
VMDRAMP'TTCYIHDDIDVAKMIQSVISRVRNIIQDQVRLQELISAQKLFHMFNQ
LPPEMNNLIDISFKYNDIPDHAQLQKQVPGIAGYGNPEQIYDITMTEMTALIPRIGHW
EFMKQSELSSKTLCLLPLKARDIANVPLVTSINDDTGNAMLAFARKIVIAFAML
HLLGENLETHDLCPQGTGVGVEGPKHCTGQALREWFIDQKTLALDAPSTHAGCA
DPMUTHGESDPNEIKMTYRNVNGVYVRKIRKICL"
78587..78592
/notes="ORF's 57 and 58"
complement(78633..78638)
/notes="ORF 59"
complement(78703..79740)
/notes="ORF 59; potential membrane protein; putative; NCBI
gi: 331269"
/codon_start=1
/number=59

```

polyA_signal

polyA_signal

CDs

```

/products="membrane protein"
/translacion="MVKGGLPVILKALKQLEGLSKAASVGTAEVIQKAYHQLTKLOAA
RLVFAVTVLIVLYIMLLIITSSVIDLVGFSPTPHRLPFPVQOQLKHFNTAQDTLIV
FVALGVNFLVLIVLYLILYLLIKLKHIEAPLSGMSYFAHPVLKYIFGINSFIFVFIIT
VFSILRTICADANTLVQDLELLSNTSFADTNE'NAAHAAVSGLLTNCPTAPHTDIAKGC
YSGVHLIAMSLESRTLRMTGTSISGLTEAGIPRMLTAVGSCWMTKEVVPVILLWFLIL
YNF LHLMMVIRALRRRLGLTIIEDEHLPLTFSYEDGEDEEGDANLLIYAIPPGSTIPG
MKKNYTPARA"
79779..80960
/notes="ORF 60; NCBI gi: 331270"
/codon_start=1
/number=60

```

CDs

```

/translacion="MMLISKARCVLELLEMP IETINGDWRKASPHVITRGTVEGVALL
HNSVVRVPGKGVSVRDANDLISFLYHDISRI SACTCEPRESMLAVILP IDKVEEY
VAKFRVVEPFTMGGFREACRAKLESAAGVTIFDRIMKDIFNRSPTLAWDGIPLAKGL
CPVLKSRGILTRGGLVGEIPKAGDHCPCKCAQYATRSRFEVTEIDLVGFDSDNT
CVLIEVTKYNSVSLP IAVLKLYNTQTYINNF LFGIMYPHLRQYKSLIAYVSPAGRV
QLFNVRSPPIITRMI SAFFPLGEYCPQMRMTAAAMTYVIKAPFAAHTVGTDNVSI
EYFSQSUKLPPNKWATDADAAREMERTEKARYRANKRANVADP'EDSP'CKELRARGKSS
T"
complement(80525..80530)
/notes="ORF 61"
complement(80932..81891)
/notes="ORF 61; NCBI gi: 331271"

```

polyA_signal

CDs

```

/codon_start=1
/number=61
/translacion="MISVARMTDSDSLVSTVDRNSDMKE IMF EKVYRGRCTHRSIP
LLDDAENFERMLCQDDLTILKEGLDFRSTPFLKSFHLTEFGLVAVTDLTETAFARI
WRANINRHVGDTADPAVEITRINHISGRTMMIPISYSGPGKQYPIILKMKHELETL
FATGKISITANHEITLPEAALLHEOHLDRLTROLP LFLYEYDGPASDPFTHVDRLGK
LRARGEVMAACVEPTPVPBSVDLMEHMSAEVCYMLLORETPTNMLMDFLHVK
TGSTTTNMLMGYNRORTISTIDILAEFLFELTAGDFTWNFALV"
81073..81078
/notes="ORF 60"

```

polyA_signal

CDs

```

81515..82819
/notes="ORF 62; NCBI gi: 331272"
/codon_start=1
/number=62
/translacion="MCRVISATVGSVSPRCTRLIARHRIAYSVRASVITSPNSVS
QNDRFNGVDRMKKSRENFRVKS'WNHNSLKAASSSGIDLCVQPKYTFSGQMF1
SEFLSVDTTRESSESVMARATEIITGAESTNKLGYCLOHELVDGFEFNRINLLVASQA
IKORVARAEKTLDELHDPVKRAMKLMGIGSDAVSRMETECERCVRDGPGERKLRFCDM
ITSVFNLMRDRGMRLEVFQVELMRGFFIGIAANQFRDDLFKYKHII LDLLGLATPDI
VAFNPECSDMTGREINRIDEYCKNYTLALAPQCCKTTIMVILLAAAMISYLDIEIV
VQAQHKTCETLYDRVELVLEHQSPWTPENRIVITNGTTFREFIYDPAYAGTTR
VHFLSSSPNSIVLHLLHSLINRDRVRCRAGVAGVCCRCA"
join(81515..82717,97459..98016,98746..99543)
/notes="putative product; putative; NCBI gi: 331210"
/codon_start=1
/number=62

```

CDs

```

/product="terminase"
/translacion="MCRVISATVGSVSPRCTRLIARHRIAYSVRASVITSPNSVS
QNDRFNGVDRMKKSRENFRVKS'WNHNSLKAASSSGIDLCVQPKYTFSGQMF1
SEFLSVDTTRESSESVMARATEIITGAESTNKLGYCLOHELVDGFEFNRINLLVASQA
IKORVARAEKTLDELHDPVKRAMKLMGIGSDAVSRMETECERCVRDGPGERKLRFCDM
ITSVFNLMRDRGMRLEVFQVELMRGFFIGIAANQFRDDLFKYKHII LDLLGLATPDI
VAFNPECSDMTGREINRIDEYCKNYTLALAPQCCKTTIMVILLAAAMISYLDIEIV
VQAQHKTCETLYDRVELVLEHQSPWTPENRIVITNGTTFREFIYDPAYAGTTR
VHFLSSSPNAAQQIPDFVLIDEAFAFNPAISLLSLP LLAVKRRKQIISHSIAKSV
TRVETIMSGDKRAFHVINQRFKCDGHAHLPQMMCP CSAFCPTHIDLANESIQLLINN
ITPGGCELELTGGGAGVLKEDVATPFSEELVRQGFHTNAPVDSLGVKRVF IAVDPTYA
NGTQSGFGVCTFAELNNGCMFLVALEEAIGEMRSILMHFTYTVVLAHVDVILGGINR
KODIP IIFLEPONTMFEDETLMSHLSGAERKFCIKLIVYRQFNKSGTTEIGLVSGN
KASWIKVHNMEADMLGRVMTYASFGELVKGLYLHNKHVLSKYIRADVTXTIIDARO
AIANPRDIEDFKKLPADSAEYIFRYLGGTEVSVAHGDDGIFKAGTELTVRIGCELMN
VSLKSVGRTKVITGKRYQGSYSRDDILFSAFILLGTSMMGRPDATDGLFIVT"
82758..84746
/notes="ORF 63; NCBI gi: 331273"
/codon_start=1
/number=63

```

CDs

```

/translacion="MYVVSVPEP LLAWCVADARDDAVIKHAIKRWICATVDIQPVTL
GAVFFTRIPSVSGRLVENSDFHMTKLP IDGAQO IDTVDRMIALAGTCGGEANVPSL
LENWFLGFEPEKTATMSALYTMESALYNHLLBKGDVIAFHIIIFERFCCQDLHLGPV
PIDVFEKWIIPDALLINLERELGSIYNLDKRNFTKESRPRLGKVLKRWLADNMIKTN
PHOIKYCLNDODDPLSVSPLPPRNALPHVP ILLMDDSSMHGEGVOYVITDYKSNETV
PPGAYCLEPQGLQVTVINFDIRKESGLDDPMCAIDNICEIFVAF LHRHVEKEFFNFN
RVOKNRIGEAVFVREALPGKTSARFTWFPAYELCFQITREAADFTGVFQELIMDE
SFFVYTVTNDGVSTRVCAIDAOPFCRNKSCRLPNSTKMEAGEPRGAFEYIKSYNALVK
SNRGYSKNNIGISRSFADPRLSGQYIYFGLARFASAEVYTNYSHASENITVEKDH
IDNAKYLKLTWIGDILKISPTTSGSVRLTPWMTDRDRECLVHNRTHHKAGVSVITVNRRI
YPKCFEHPDPPAHVPIEGGFLDWVEVAGKPRARVFCGQ"
84664..86208
/notes="ORF 64; NCBI gi: 331274"
/codon_start=1
/number=64

```

CDs

```

/translacion="MSSPRGASSMSRSPVNLFEPSDVSVDSEWFKYKRLGALRSAT
LEGEDLEGQIPQELLNVLVLYERKMKSTKSELSGFVYIALMTLGLIFDORSIKRLVDE

```


SDTKLYTVREGECEPTTAVIPDPTHSHVVPATTPAPEAGCKLMSDPLELIRKGGK
 RVIMDSVGVASGLVAGSPSERVMKRSERPLDSSAVKRVSKVTSVTDQECTAM
 ESSIPGVSTERAPTITYGKSAQTTAGVLTEGGVAPYPLIMVDVGERDLPIDLPNQIIT
 GRVIESIGLEAGDIAARKEVKKAVEALAVSYRPLMQLSLGIFPDEGHIYSVLEXP
 TGDILMSQDTLRVLSITFTFPMPGKSPAGTKMAIVHHIFRDVVKLVQNLQIRVTPV
 FKQDVYVWEHGLTIAPNSVRKDTICQNLIIISLSQVTFQHTTKLP IKPEVRAT
 LSPDVLKLDMPRPSPSEVTEMACRIGVNGDIGHPTIOFPLDEKIHMGSEVRLPDV
 LLIKPKAAPSYFSPSMEVIRVLETPGRSDESDWVPASALISHATLAASORGLVAA
 HASVGAQAALSITGGAFCATPSEGDAITSAIMKMIQELSNTSIPTSLSLCEGLSSM
 RGP AVATPPACAPPSYSGEGLVDFALEKWTGPADVECTAPVVEIP IVGGDI IRQ
 SNA LAGECHELITETSPGNVATGETTPATPTIINPSPHVPVMPPPANCKRTKS
 VKRIEASKEGGERDATGSSMOTVRGTGGLVTPKQSLIRGDELGLMKWVFEPAL
 DTRTSYLGVRHTIIVQMLQHGSKMKRCRDPPTFAITDADVP ILTTPDATRRRI
 CVDNLPSTPWFVWGPNGSEPPVTEHRTSYELRGDIQAAHQCNILYLPSEIYAY
 SPRVMIAQHTDGAHQ IGGGFPPIRVYVYINNMKIRPSKHIPYIYVIEPHSVGYKM
 NYAYLITPMGHVLTTCVAGYSKYLKALVRLT LHAHVSAGLKI GYMPHKNVWKTDAL
 GSDLOTYIFDVVQIILENVFMPGSGGRNAPVIPPWTSMKTDNGCPPNLITIASDI
 FCLGRLLTVFCAQDQLTSQVALIQIMCAKKEERPCVTDLLVNFVSPPEEPFYPVDI
 GEMPVKITNTKKAERSNTAYVRFVPQAGE"

repeat_unit
 polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal
 CDS

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

polyA_signal

MKCSGLLKRCVREXIPVSDRGFLDAYTQTAPGIVFTGSSLVTDSTLMMARFNH
 RQHVLPLEENFGLERARGDVTVDIRTDDEAVARGLSDMYERDAVGHIVSKPQRTT
 GEMULDELMWARTLSHGNLVRELGVAACEETRGCKITVKMVELGLONLDEFIAEPHMIA
 PKAIVROYTICKLERDITSLGLEYLHMHGVHGDGFRGAKVIVFP TADDETFKNLGEES
 TSIGLAPLNRAPYHRAPEFTEGLIGREADMWALMAYRAHTCKDFKKGDPSEFY
 PARELMDNDALDSRRTGRDGPVPEGFCITRLEKNATVGDGFI GLMCKCLALYPOFR
 RTARELIVPRYATLHRTSGPAPPVIGAAPAVHVSNAVERSRGVRSPAMPD
 RLEARTSGKNANAVGDFTPSFYHREGGLDTRPKRAITLSRGTLASVYGTENRL
 TTEIHRSGADIVLHKGVRPDI SRANLEYMKVNGPEMYITIDHYTGVGVNHL
 YVGYITDDIVASAGLSCP LRVDILKGFESTLVRDLMGLVTRGLAYTLVRVHLK
 SASIEAGPHAVGLVFNWPSNRNFCVSCDKVLELDEPESQLITCSR"
 complement (110249..110254)
 /notes="ORF 75"
 complement (110271..111542)
 /notes="ORF 75; NCBI gi: 331283"
 /codon_start=1
 /number=75
 /translation="WNTLKSSQCVHCILEAYPPGAGPVSALDERWIRIFMPGGETLM
 IDRRIVYEPSPCTYHTGLDVLGAHIVPNNSFDRDVTGRDLSDFRSCDFMEDVHPW
 FRVLEITTDRLDGPGRGETVINFGRGVPGVYLCGSHGVATSCITMGPVQGMPE
 TISITINPFGGFWSLWLKPACTGSKSPFANDLTRYLIRTSFGSTAGIDRETFESP II
 LYGGVWLECVFPHLPTLTVAKESTECALDPSGIWVIMPALPVVRHVTLYVDR
 VROLVCTLAANFEDIFVHRVDRRVTHRWEI PKNAPRCKLSFTVTGHRPEVLAAMCG
 GMRGLTNPOMGFDITDLKVEPAPGVKLDGPPRLHPEFLAAMAKIEFLVAP IESEL
 RGMTSLHAWTITVVALVGRFGILITALESL"
 complement (111526..111531)
 /notes="ORF's 76 and 77"
 complement (111560..111565)
 /notes="ORF's 76 and 77"
 complement (111597..112322)
 /notes="ORF 76; related sequence to ORF 77; NCBI gi:
 331284"
 /codon_start=1
 /number=76
 /translation="MDRHRYLSILTPELVAGVPAAYELFMTERGSGIALIPPVDRSD
 ESRVESIIIRAHGRPIIAFAGRMGSGKHACDYCFERFAGAIKLHIFESGLRIGLYGR
 AIDKSDRSVLQVAGELGRRLRQFWMQTEEDLLRLILDRDPRVITGVYKTDVRY
 LATLGVRTFIIRRPGLVLAGTPEESDPSERDLDVYHECAELINRGTVGDFALLAETA
 RDISLSLVPTDTGRECPRIEP"
 complement (112497..113183)
 /notes="ORF 77; related sequence to ORF 76; NCBI gi:
 331285"
 /codon_start=1
 /number=77
 /translation="MARKGEPPSYKALTPEIVSAVPEAYPLFVYSEKHMSSGFP
 ENSNRSTGRYDWMVYDLARTKGTTLTAFAGRMGSGKHACDYCSIKIGGEKVHVDAG
 LMRATKLGRRTEKPRDRPFLQALGDUGRRLDANFWWRDAATEKIVAPMWSGVNVFIT
 GVAFPEALAGLGVRTLVRPRLITGSPMERHRSSETALDEYRDCPEILNDSTIAA
 FEDRVSTELP"
 113341..114543
 /notes="ORF 78; potential Zn-binding protein; putative;
 NCBI gi: 331286"
 /codon_start=1
 /number=78
 /product="Zn-binding protein"
 /translation="MELSUYRELPRKRAVQAPRTQPPPKVHREDTGVIPGLVRESS
 VRPDSITNFRLLAAVKFNPEWMECGGCGCECKTRPALMNCDCDCTAKYPIICLGCK
 PRVRDHPCFNRRTVCSGCTIGVAVVWKLGGCGATCLDCSLGADWPKRSRTGTGCVG
 CAOTLGVHYRPMKI IKNTPIYPAETPIPQVQRIHISARILEIRCOQLSRSEIYSLFG
 DQGSMMYEPAPVAPHLAPFMIGSDISKAFADDTILPFCRAMCAANGVPGWRGQQA
 PVLPFGMDMCLGCVRAQEQITIQVINGTIWEDAP IAGTYHMSWILDPITVTVSPVD
 ISALDTEAVEFCNGLGSHPEKYRLSDITCRIRWTAGGFLDPESLWPEKELKNTA
 SAHSNNSH"
 114588..115133

CDS

SDTKLYTVREGECEPTTAVIPDPTHSHVVPATTPAPEAGCKLMSDPLELIRKGGK
 RVIMDSVGVASGLVAGSPSERVMKRSERPLDSSAVKRVSKVTSVTDQECTAM
 ESSIPGVSTERAPTITYGKSAQTTAGVLTEGGVAPYPLIMVDVGERDLPIDLPNQIIT
 GRVIESIGLEAGDIAARKEVKKAVEALAVSYRPLMQLSLGIFPDEGHIYSVLEXP
 TGDILMSQDTLRVLSITFTFPMPGKSPAGTKMAIVHHIFRDVVKLVQNLQIRVTPV
 FKQDVYVWEHGLTIAPNSVRKDTICQNLIIISLSQVTFQHTTKLP IKPEVRAT
 LSPDVLKLDMPRPSPSEVTEMACRIGVNGDIGHPTIOFPLDEKIHMGSEVRLPDV
 LLIKPKAAPSYFSPSMEVIRVLETPGRSDESDWVPASALISHATLAASORGLVAA
 HASVGAQAALSITGGAFCATPSEGDAITSAIMKMIQELSNTSIPTSLSLCEGLSSM
 RGP AVATPPACAPPSYSGEGLVDFALEKWTGPADVECTAPVVEIP IVGGDI IRQ
 SNA LAGECHELITETSPGNVATGETTPATPTIINPSPHVPVMPPPANCKRTKS
 VKRIEASKEGGERDATGSSMOTVRGTGGLVTPKQSLIRGDELGLMKWVFEPAL
 DTRTSYLGVRHTIIVQMLQHGSKMKRCRDPPTFAITDADVP ILTTPDATRRRI
 CVDNLPSTPWFVWGPNGSEPPVTEHRTSYELRGDIQAAHQCNILYLPSEIYAY
 SPRVMIAQHTDGAHQ IGGGFPPIRVYVYINNMKIRPSKHIPYIYVIEPHSVGYKM
 NYAYLITPMGHVLTTCVAGYSKYLKALVRLT LHAHVSAGLKI GYMPHKNVWKTDAL
 GSDLOTYIFDVVQIILENVFMPGSGGRNAPVIPPWTSMKTDNGCPPNLITIASDI
 FCLGRLLTVFCAQDQLTSQVALIQIMCAKKEERPCVTDLLVNFVSPPEEPFYPVDI
 GEMPVKITNTKKAERSNTAYVRFVPQAGE"

CDS

CDS

CDS

CDS

CDS

CDS

MKCSGLLKRCVREXIPVSDRGFLDAYTQTAPGIVFTGSSLVTDSTLMMARFNH
 RQHVLPLEENFGLERARGDVTVDIRTDDEAVARGLSDMYERDAVGHIVSKPQRTT
 GEMULDELMWARTLSHGNLVRELGVAACEETRGCKITVKMVELGLONLDEFIAEPHMIA
 PKAIVROYTICKLERDITSLGLEYLHMHGVHGDGFRGAKVIVFP TADDETFKNLGEES
 TSIGLAPLNRAPYHRAPEFTEGLIGREADMWALMAYRAHTCKDFKKGDPSEFY
 PARELMDNDALDSRRTGRDGPVPEGFCITRLEKNATVGDGFI GLMCKCLALYPOFR
 RTARELIVPRYATLHRTSGPAPPVIGAAPAVHVSNAVERSRGVRSPAMPD
 RLEARTSGKNANAVGDFTPSFYHREGGLDTRPKRAITLSRGTLASVYGTENRL
 TTEIHRSGADIVLHKGVRPDI SRANLEYMKVNGPEMYITIDHYTGVGVNHL
 YVGYITDDIVASAGLSCP LRVDILKGFESTLVRDLMGLVTRGLAYTLVRVHLK
 SASIEAGPHAVGLVFNWPSNRNFCVSCDKVLELDEPESQLITCSR"
 complement (110249..110254)
 /notes="ORF 75"
 complement (110271..111542)
 /notes="ORF 75; NCBI gi: 331283"
 /codon_start=1
 /number=75
 /translation="WNTLKSSQCVHCILEAYPPGAGPVSALDERWIRIFMPGGETLM
 IDRRIVYEPSPCTYHTGLDVLGAHIVPNNSFDRDVTGRDLSDFRSCDFMEDVHPW
 FRVLEITTDRLDGPGRGETVINFGRGVPGVYLCGSHGVATSCITMGPVQGMPE
 TISITINPFGGFWSLWLKPACTGSKSPFANDLTRYLIRTSFGSTAGIDRETFESP II
 LYGGVWLECVFPHLPTLTVAKESTECALDPSGIWVIMPALPVVRHVTLYVDR
 VROLVCTLAANFEDIFVHRVDRRVTHRWEI PKNAPRCKLSFTVTGHRPEVLAAMCG
 GMRGLTNPOMGFDITDLKVEPAPGVKLDGPPRLHPEFLAAMAKIEFLVAP IESEL
 RGMTSLHAWTITVVALVGRFGILITALESL"
 complement (111526..111531)
 /notes="ORF's 76 and 77"
 complement (111560..111565)
 /notes="ORF's 76 and 77"
 complement (111597..112322)
 /notes="ORF 76; related sequence to ORF 77; NCBI gi:
 331284"
 /codon_start=1
 /number=76
 /translation="MDRHRYLSILTPELVAGVPAAYELFMTERGSGIALIPPVDRSD
 ESRVESIIIRAHGRPIIAFAGRMGSGKHACDYCFERFAGAIKLHIFESGLRIGLYGR
 AIDKSDRSVLQVAGELGRRLRQFWMQTEEDLLRLILDRDPRVITGVYKTDVRY
 LATLGVRTFIIRRPGLVLAGTPEESDPSERDLDVYHECAELINRGTVGDFALLAETA
 RDISLSLVPTDTGRECPRIEP"
 complement (112497..113183)
 /notes="ORF 77; related sequence to ORF 76; NCBI gi:
 331285"
 /codon_start=1
 /number=77
 /translation="MARKGEPPSYKALTPEIVSAVPEAYPLFVYSEKHMSSGFP
 ENSNRSTGRYDWMVYDLARTKGTTLTAFAGRMGSGKHACDYCSIKIGGEKVHVDAG
 LMRATKLGRRTEKPRDRPFLQALGDUGRRLDANFWWRDAATEKIVAPMWSGVNVFIT
 GVAFPEALAGLGVRTLVRPRLITGSPMERHRSSETALDEYRDCPEILNDSTIAA
 FEDRVSTELP"
 113341..114543
 /notes="ORF 78; potential Zn-binding protein; putative;
 NCBI gi: 331286"
 /codon_start=1
 /number=78
 /product="Zn-binding protein"
 /translation="MELSUYRELPRKRAVQAPRTQPPPKVHREDTGVIPGLVRESS
 VRPDSITNFRLLAAVKFNPEWMECGGCGCECKTRPALMNCDCDCTAKYPIICLGCK
 PRVRDHPCFNRRTVCSGCTIGVAVVWKLGGCGATCLDCSLGADWPKRSRTGTGCVG
 CAOTLGVHYRPMKI IKNTPIYPAETPIPQVQRIHISARILEIRCOQLSRSEIYSLFG
 DQGSMMYEPAPVAPHLAPFMIGSDISKAFADDTILPFCRAMCAANGVPGWRGQQA
 PVLPFGMDMCLGCVRAQEQITIQVINGTIWEDAP IAGTYHMSWILDPITVTVSPVD
 ISALDTEAVEFCNGLGSHPEKYRLSDITCRIRWTAGGFLDPESLWPEKELKNTA
 SAHSNNSH"
 114588..115133

CDS

```

/notes="ORF 79; NCBI gi: 331287"
/codon_start=1
/number=79
/translation="MDEFGSLVTDGERVSNTRDGRFMSAARPGMLKPSGRADGGV
WFCDDVTDYDITVAHGTLQDDNLFVMTRLFPGGGLIPDDVKSDRHVCPDCGVKIVR
FTLVDDFLDEHYHSGGCPRIROFLQORVSELYELTQFKGDRPYPARRKWDLSRD
SAVLKAVAYASPKDQNSVQTG"
115168..115242
115246..115251
/notes="ORF" s 78 and 79"
115291..115447
115567..115572
/notes="ORF" s 78 and 79"
115733..115943
complement(116137..116142)
/notes="ORF 1"
complement(116253..116258)
/notes="ORF 1"
116279..116418
complement(116412..119000)
/notes="ORF 1; Identical sequence as found in span
742-3330; NCBI gi: 331289"
/codon_start=1
/number=1
/translation="MDGKIEITAAVASLGGTTDLSTVWVNFDDGMDQASGWDIDV
HPVVEGASVDAMDVAESDEKDEEPTGRDEPEVESEADDEVEDAIPAAVPT
IPLSTTIRGEGADLGLKNLTPADLAGNLGKYAATLFSEKTLPTLAAVVEQKALDGG
SSVAKSVTLISSELGNALGIFSRPLTREONMAAVMLIMKVLGRTALPLPLTKTDEGV
LSAIVLFSKARKVILVTHSDITKFIENDDGTIPTLGVDVAPLAKYVDPRLKLT
LVLIKLNNGTVKVRKSKTIVMERIPVECFLEFRAPPTSESSUNIIYATLDFVR
LIKGVTSQSVYVYVVKPSIGDDFTKNTKTSITGLDAPLTVKLSFRKPTNATSSGFK
RMLIKCKLARGRIELGPKSTLRSGHEESLEEVTGCIIRSMASLTRIASRAMSPK
NLIKVNDPVPMTSILNEIPDEVHRTPLIPWLLFAASLRLDIVAVKEWLESTEGCDPR
SIDVAKITVHYVLELRLLSDAGIAKGLSKLLITANMLITEMSGIVMYKGSYGNPT
VLSTISDNVNPILPISWIEIPLWSTSGIVKVKIVHTNSSILVSRMLLSIDVTEVAK
VHAIAKAVRAEQKASKESAKGADAAKPKRKEAGETETPKKKQKEEKSEKPKYK
KAESTGTAKVTKEKTEKTPQTKSTNKNKVSVAEAGTAVPDEANGKKSTQVK
RKESAGQHTTEKKRTAAKKTVDPRSGHRPSKREYRSQEFVSSDDSDSEEVISKPR
VTTENMKSTKATLASNMDDDDGDMTCAGQFDIGCGSGFLTSTAGNRKSNKQS
KTLILS"
119172..120395
/notes="ORF 2; Identical sequence as found in span
3502-4725; NCBI gi: 331289"
/codon_start=1
/number=2
/translation="MTCGRHSPHQTSPPRPVEKMPRFQEHGASHRYTPMCKRPAAP
ANTGSHRGALLGRNGDPERKEETAVSKLILKSRGYIPEARTALIEWKEGARGGF
EVMCGSITLLPGHDSGRAIDDIFRDMRSVRGEVASPIDELLFRVGNLLEVYVETTG
KEVPKAGKGVQVQSVTVTGIPKMIHVCLDRSLGVGKYSFAKLWKTIEELLGTGCG
TTPHQCMTFRTIHSDAIOASFDGLNMIIEFLGGRVTRGVRSPGERRVFTHRLCGDS
FNAFLAATVHDDTVYSRAVFEARVTVQTSYPAMYTFGDMTPPAPNNVGMVTTGSGMC
YDIRVRLDLYNDPNKMSVTPAAAAPPVGPKEGHEELEADPHKPSSTIGSPAPNICALY
TSGDSVELHAPLLPL"
120556..121464
/notes="ORF 3; Identical sequence as found in span
4886-5794; NCBI gi: 331290"
/codon_start=1
/number=3
/translation="MAFSTSDKRIAEILLGMSPATLSVRVFGVDLPHPIDWSDDEG
DDTPTGWSERALGVTVARPHVYIGPVGICAVDDSDTSDMSPASPVYGAERAGG
LPQAPALLVRRVSGNEMRVSESRRIEIEIGRELARICEIQWEDVAGAPVGTGNS
PWSVPGSLLLPKSGSPTYWNPSPSPSDTTCELPISLPSLVSSPVNPAPNYSFVSD
APLNTFSPNNMITYTHPGARLLDTPPTFLDADGDWAVDWLYAPQLGQQLIESGSS
PKRPLDDPDEVPSSKSGPSRRALLN"

```

```

repeat_unit
121338..121626
121754..121911
121788..121793
/notes="ORF" s 2 and 3"
121828..121833
/notes="ORF" s 2 and 3"
121864..121869
/notes="ORF" s 2 and 3"
121904..121909
/notes="ORF" s 2 and 3"
121937..121942
/notes="ORF" s 2 and 3"
122183..122956
123630..124169
/notes="ORF 4; Identical sequence as found in span
7960-8499; NCBI gi: 331291"
/codon_start=1
/number=4
/translation="MASIDVGETIGIEQVPEVKEFVCHGCGCPSPFVTARGAGSVSL
RVVHYVSPEREDPESPTRGVDVDCACSEPTTPRPERGRAIEEMGKIIVILVSVCP
LRPALQHRVAGCAPARRHRSDSVARRAFEPWRGRASRPTVLTYPASGSDSDVDTRTFC
CGCPGHPFVDCMCDQDML"
124227..124232
/notes="ORF 4"
124312..124317
/notes="ORF 4"
124455..125141
/notes="ORF 5; Identical sequence as found in span
8785-9471; NCBI gi: 331292"
/codon_start=1
/number=5
/product="thymidine kinase"
/translation="MALREGPTPARELPEGLVFCVEGNIGCGKSTLVKALMERVAGSG
VNVVEPVDQWVNHKNYIELSYDTPGYAVFQNLVFDYVNVQVQNLQNDIMERSP
MSATRVFCVNGSRGVPATALPCMAARGEAVMRTIATRPVYVLELPPEELARRMR
RDRTEAGVGGLDYLRLLHRYEAWLSSAEDVERVDASRSEEVDRVRIELLCRHRL
RAPLTRKSQL"
125205..125621
/notes="potential product; ORF 6; Identical sequence as
found in span 9535-9951; potential membrane protein;
putative; NCBI gi: 331293"
/codon_start=1
/number=6
/product="membrane protein"
/translation="MNSLTIIIFLSGLTAYHAVIADGTGSSSVTAGDSGVVWLMVIG
ALLTLMTIPIIGLGIYVTRASIEEMRGLMQIHLRLITGDRSNRGDELGAGAS
LLTSSQPPSYAEALLMEPEVPPQQEGVPLEAEIRV"
125713..125718
/notes="ORF" s 5 and 6"
125918..125937
125968..125996
126090..126284
126533..126955
/notes="potential product; ORF 7; Identical sequence as
found in span 10863-11285; potential membrane protein;
putative; NCBI gi: 331294"
/codon_start=1
/number=7
/product="membrane protein"
/translation="MAVILIERAAEFVAPGEARVGPVILAEVYRALTSDHEMRAFYET
CAVTSFADMLIIWLVHARHPGCTTGTDAHTQTEGSTTGTDAHTQTEGSRDQGS
MTEADDLITRPPLGCHGRQIPVIRRMVLDRLRIDYSL"
126729..126822
127169..127174
repeat_unit
polyA_signal

```

```

repeat_unit
repeat_unit
CDS
/notes="ORF 7"
127328..127421
127437..127568
128132..128830
/notes="potential product; ORF 8; Identical sequence as
found in span 12462-13160; potential membrane protein;
putative; NCBI gi: 331295"
/codon_start=1
/number=8
/product="membrane protein"
/translation="MKCPARVDIFKREIKTRELEPTKKNVHGLTGPISLPPHPVS
IRFQMSVRLRSVIGFTLLVSSGSSAASIP SAPTADATRESPTCEPHDRALST
ETPTEPSRDGSGTEVLHVTPGVRPRDRPILERLAEILAEHLSHOLLHPTGTPR
EDEDVFAALAAEIALGSAVDRVMKATLSCHLVTSVTFAGVALWIVARHGHR
VIPHERSDWSPGL"
128867..129343
/notes="potential product; ORF 9; Identical sequence as
found in span 13197-13673; potential Zn-binding protein;
related sequence to ORF 10, ORF 11; putative; NCBI gi:
331296"
/codon_start=1
/number=9
/product="Zn-binding protein"
/translation="MATPKVRRPRDRLVPPKVDRIKRVIFACGHRKRCIGYRPLG
DSGNGCPGCAARDVADPDIYSGPLPPVCGRAWGPTVREACGHVTCNACETECA
VDRLICGGRRILVAICDPYPPGPRMRGPRTPREAHEAQRSGSSEDACTAP"
129384..129403
129446..129451
/notes="ORF's 8 and 9"
129746..130201
/notes="potential product; ORF 10; Identical sequence as
found in span 14076-14531; potential membrane protein;
related sequence to ORF 9, ORF 11; putative; NCBI gi:
331297"
/codon_start=1
/number=10
/product="membrane protein"
/translation="MTVKGLCLAFGVTLIVGVVWVAMGVALSKRAPGDDLVTLAL
NELAGDDVVRGLSDAANLTGLIRGLILDHLSRNSGSLVAADALVERVLTRILDI
FVFNGLVAENCFWLSIFIMYTCAITVRSKKVAPAIHSPPTMSPYI"
130209..131054
/notes="potential product; ORF 11; Identical sequence as
found in span 14539-15384; potential Zn-binding protein;
related sequence to ORF 9, ORF 10; putative; NCBI gi:
331298"
/codon_start=1
/number=11
/product="Zn-binding protein"
/translation="MRCIRAPADDDGYPTIIMSPVPCFCQGEAVFPNSRCKHY
FCFTPREPCPGSGGKRLIPNEYLIYALTANPPAPMGRTAGVIMPGMGHVE
PRVVLLEDLLTAVIITVGLAVETRGAGEAVRARGKEGAIALPLLDLDLVELGG
AIEAAGCKVAGGFLVRLTYELVRYHDITKTFPVNAPRFGSLGALKELLSRRIFPC
YFGSGTPRDYDGLLGHINACEIRKCCDPPGCGFNFRAFLNRKNRHKRTGGATGSPFID
PPTGH"
131061..131960
/notes="potential product; ORF 12; Identical sequence as
found in span 15391-16290; potential Zn-binding protein;
putative; NCBI gi: 331299"
/codon_start=1
/number=12
/product="Zn-binding protein"
/translation="MTSPREITVRLSCGPIRLVPDGEFCFGAYCCDCDERAVFA
PDHLITVLSKVPAGGLDAITEKSWKLAKAIVTCGTGCVLDVQPVIERHCGHCRFCF
RDKCAEGGLMARTAVRPLPDLRLIQESLFCGHTRGLEYSDKFTLAELMGPGQKVELLY

```

```

DGLGNVFDLVITGISAVIADRAEKPIDLVDIVLPDMSVPVCTIPYRLRLIKANA
EMGVHWDSDIVITGAMNLTVEYGAVPGLDRSLIEFLRYLVPFGDAETFPALTET
LSLIAECTWGDRCRCYDACS"
131999..132004
/notes="ORF's 10, 11 and 12"
132024..132272
/notes="ORF 13; Identical sequence as found in span
16354-16602; NCBI gi: 331300"
/codon_start=1
/number=13
/translation="MAFLIPFLCNCNPMSLLGGGCDLISCCRCGGGQPMARQPIY
PYGSPMGARVYPPVPAQPPVRGVRVPPQGERPVDGLR"
132335..132386
repeat_unit
polyA_signal
132410..132415
/notes="ORF 13"
132414..132419
polyA_signal
132419..132419
/notes="ORF 13"
132717..133026
repeat_unit
polyA_signal
complement(132746..132751)
/notes="ORF 14"
complement(132798..132803)
/notes="ORF 14"
complement(132849..132854)
polyA_signal
/notes="ORF 14"
complement(132901..132906)
polyA_signal
/notes="ORF 14"
complement(132968..132973)
polyA_signal
/notes="ORF 14"
complement(133016..134191)
/notes="ORF 14; Identical sequence as found in span
17346-18521; related sequence to ORF 15, ORF 16; NCBI gi:
331301"
/codon_start=1
/number=14
/product="protein kinase"
/translation="MAEKLIPSCRGOWFIPARVIREGFIPARVAVDVGKLTIGPLT
VATVMDVYVYHNENATDHLRGNPFPWEEPQDVLDFPGVACVNEATFLTKKYG
SYLHLPVLRHTRGNTFFGSPLSATGMALIKTFHGTOAPDRPGVPMVEYAMETA
HFLDAPKLIEWEVSCTRENKSAQWTFSEGLYSGMEGYFRERATVDICATILAD
LTGKLLALIRKGIYHCDLKSENIIMSRSGPKLIDFEHSGCETMTSFYWPRTFF
WNPDIGTEYAPPPERSRGRGNAGVGVIFQIGLIALINIMVERVFNHHTWIKGGY
RAHLVKVIRKAGTDLRGGARTLARVDELIGLVARCLERDPANRPSLETFLVDEFSKI"
BASE COUNT 28727 a 37766 c 37707 g 30026 t
ORIGIN

DB 71; Score 24; Match 78.6%; QryMatch 1.6%; Pred. No. 1.46e-01;
Matches 33; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 42146 ggaacttcaccagaagtcgaagtcacattacacaggagc 42187
||||| ||||||| |||| ||| ||| ||| |||
Cp 679 GGAAGCTTCTACGACGACGACGACGACCTTCCCGAAGC 638

RESULT 11
LOCUS TRFCE1 3940 bp DNA INV 15-SEP-1990
DEFINITION C.fasciculata retrotransposable element I (CRE 1) gene complete
cds.
ACCESSION M33009
KEYWORDS integrase; retrotransposable element I; reverse transcriptase.
SOURCE C.fasciculata.
ORGANISM Crithidia fasciculata
Eukaryota; Animalia; Protozoa; Sarcosomastigophora; Mastigophora;
Kinetoplastida; Trypanosomatina; Trypanosomatidae.
REFERENCE 1 (bases 1 to 3940)

```

AUTHORS Gabriel, A., Yen, T.J., Schwartz, D.C., Smith, C.L., Boeke, J.D.,
Sollner-Webb, B. and Cleveland, D.W.
TITLE Rapidly rearranging retrotransposon within the minixon gene locus
of *Crithidia fasciculata*
JOURNAL Mol. Cell. Biol. 10, 615-624 (1990)
MEDLINE 90136575
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by A.Gabriel, 21-MAR-1990.

FEATURES NCBI gi: 162492
source Location/Qualifiers
1..3940
/organism="Crithidia fasciculata"
repeat_region 1..415
/note="mini-exon gene repeat"
misc_feature 386..415
/note="target site duplication"
misc_feature 415
/note="mini-exon splice junction"
misc_feature 3899..3927
/note="target site duplication"
repeat_region 3899..3940
/note="mini-exon gene repeat"
BASE COUNT 890 a 1032 c 1330 g 688 t
ORIGIN

DB 31; Score 24; Match 67.6%; QryMatch 1.6%; Pred. No. 1.46e-01;
Matches 46; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Db 2493 atggtgtggggccgaagcaaccgggtggggtgagctgggggtgagactgcgggtgtgtg 2552
|||||
|||||

Cp 1174 ATGGTGGTGGAGGCTCAGGAACAGGGGTGGAGTGGGTCGATCCATGGTGGTG 1115

Db 2553 gccgagcc 2560
|||||
Cp 1114 GCGAGGC 1107

RESULT 12
LOCUS MUSOF1 1514 bp mRNA ROD 13-JAN-1992
DEFINITION Mouse mRNA for OSF-1.
ACCESSION D90225
KEYWORDS OSF-1.
SOURCE Mouse (new born; strain C57BL/6) calvaria osteoblast cell, cell
line MC3T3-E1, cDNA to mRNA, clone pMC031.

ORGANISM Mus musculus
Eukaryota; Animalia; Chordata; Mammalia; Rodentia; Muridae.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Tezuka, K., Takeshita, S., Hakeda, Y., Kumegawa, M., Kikuno, R. and
Hashimoto-G, T.

TITLE Isolation of mouse and human cDNA clones encoding a protein
expressed specifically in osteoblast and brain tissues
JOURNAL Biochem. Biophys. Res. Commun. 173, 246-251 (1990)
MEDLINE 91076859
COMMENT These data kindly submitted in computer readable form by: Reiko
Kikuno

Pharma Research Labs.
Hoechst Japan Ltd.
1-3-2 Minamidai
Kawagoe, Saitama 350
Japan
Phone: 492-43-6149
Fax: 492-45-6036.

FEATURES NCBI gi: 220552
source Location/Qualifiers
1..1514
/organism="Mus musculus"
CDS 215..721
/gene="osf-1"
/note="NCBI gi: 220553"
/codon_start=1
/product="OSF-1"
/translation="MSSQQYQQRKFAAFILAFILAAVDTAEGKKEKKYKK
SDCGEMQWCVPTSDCGLGTRGTRGAECKQTKTORCKIPCNKKQKQAECKYQ
FOAGCEDLNTALKTRTGSIKRALHNADQKTVTISKPCGKLTGPKPAESKKKKEG
KKQEKMLD"

polya_signal 1476..1481
BASE COUNT 484 a 267 c 350 g 413 t
ORIGIN

DB 57; Score 24; Match 71.4%; QryMatch 1.6%; Pred. No. 1.46e-01;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 135 cctggagctgcgaacactgcgccttgcacaaaggcagcagctagtca 190
|||||
|||||

Qy 980 COTGGAGCTCAGACATCTGCCCGGAGCTTTCACAAAGGCTCCCTGGCATTCA 1035

RESULT 13
LOCUS HUMPKD1GEN 53522 bp DNA PRI 28-APR-1995
DEFINITION Homo sapiens polycystic kidney disease-associated protein (PKD1)
gene, complete cds.

ACCESSION L39891
KEYWORDS polycystic kidney disease.

SOURCE Homo sapiens (clone: cggg10.2, cDEB11) DNA.

ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.

REFERENCE 1 (sites)
AUTHORS Ward, C.J., Peral, B., Hughes, J., Thomas, S., Gamble, V.,
MacCarthy, A.B., Sloane-Stanley, J., Buckle, V.J., Kearney, L.,
Higgs, D.R., Ratcliffe, P.J., Harris, P.C., Roelfsema, J.H.,
Spruit, L.L., Sarris, J.J., Dauwerse, H.G., Peters, D.J.M.,
Breuning, M.H., Nellist, M., Brook-Carter, P.T., Maheshwar, M.M.,
Cordeiro, I., Santos, H., Cabral, P., Sampson, J.R., Janssen, B.,
Hesseling-Janssen, A.L.W., van den Ouweland, A.M.W., Eussen, B.,
Verhoeve, S., Lindhout, D. and Halley, D.J.J.

TITLE The polycystic kidney disease 1 gene encodes a 14 kb transcript and
lies within a duplicated region on chromosome 16. The European
Polycystic Kidney Disease Consortium [published erratum appears in
Cell 1994 Aug 26;78(4):following 724]

JOURNAL Cell 77 (6), 881-894 (1994)
MEDLINE 94273192
REFERENCE 2 (bases 1 to 53522)

AUTHORS Burn, T.C., Connors, T.D., Dackowski, W.R., Petry, L.R., Van Raay, T.J.,
Milholland, J., Venet, M., Miller, G., Hakim, R.M., Doggett, N.A.,
Landes, G.M., Klinger, K.W., Qian, F., Onuchic, L.F., Watnick, T. and
Germino, G.G.

TITLE Analysis of the genomic sequence for the autosomal dominant
polycystic kidney disease (PKD1) gene predicts the presence of a
leucine-rich repeat

JOURNAL Hum. Mol. Genet. 4 (4), 575-582 (1995)
COMMENT NCBI gi: 790818

FEATURES Location/Qualifiers
source 1..53522
/organism="Homo sapiens"
/clone="cggg10.2, cDEB11"


```

intron      15818..19865
            /gene="PKD1"
            /map="16p13.3"
            /number=2
repeat_region complement(17907..18198)
repeat_region /rpt_family="Alu"
               complement(18260..18516)
repeat_region /rpt_family="Alu"
               complement(18574..18843)
repeat_region /rpt_family="Alu"
               complement(19064..19350)
exon         /rpt_family="Alu"
               19866..19974
            /gene="PKD1"
            /map="16p13.3"
            /number=3
intron       19975..20095
            /gene="PKD1"
            /map="16p13.3"
            /number=3
exon         20096..20167
            /gene="PKD1"
            /map="16p13.3"
            /number=4
intron       20168..20435
            /gene="PKD1"
            /map="16p13.3"
            /number=4
exon         20436..20605
            /gene="PKD1"
            /map="16p13.3"
            /number=5
intron       20606..21109
            /gene="PKD1"
            /map="16p13.3"
            /number=5
exon         21110..21490
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=6
misc_feature 21174..21459
            /gene="PKD1"
            /map="16p13.3"
            /note="CpG island; putative"
intron       21491..21608
            /gene="PKD1"
            /map="16p13.3"
            /number=6
exon         21609..21792
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=7
intron       21793..22227
            /gene="PKD1"
            /map="16p13.3"
            /number=7
exon         22228..22448
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=8
intron       22449..22636

```

```

exon         22637..22752
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=9
intron       22753..23117
            /gene="PKD1"
            /map="16p13.3"
            /number=9
exon         23118..23289
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=10
intron       23290..23655
            /gene="PKD1"
            /map="16p13.3"
            /number=10
exon         23656..23815
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=11
intron       23816..24966
            /gene="PKD1"
            /map="16p13.3"
            /number=11
exon         24967..25111
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=12
intron       25112..25988
            /gene="PKD1"
            /map="16p13.3"
            /number=12
exon         25989..26120
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=13
intron       26121..26317
            /gene="PKD1"
            /map="16p13.3"
            /number=13
exon         26318..26493
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=14
intron       26494..26807
            /gene="PKD1"
            /map="16p13.3"
            /number=14
exon         26808..26941
            /gene="PKD1"
            /map="16p13.3"
            /note="putative"
            /number=15
intron       26942..27406
            /gene="PKD1"

```

```

/misc_feature
27407..28254
/notes="CpG island; putative"
/number=15
/feature="PKD1"
/number=16
/feature="putative"
27526..27627
/feature="PKD1"
/number=16
/feature="PKD1"
27628..28683
/feature="PKD1"
/number=16
/feature="PKD1"
28608..28993
/notes="CpG island; putative"
/number=17
/feature="PKD1"
28684..29125
/feature="PKD1"
/number=17
/feature="PKD1"
29126..29875
/feature="PKD1"
/number=18
/feature="putative"
29857..30825
/notes="CpG island; putative"
/number=18
/feature="PKD1"
30419..31024
/feature="PKD1"
/number=18
/feature="PKD1"
31025..31243
/feature="PKD1"
/number=19
/feature="PKD1"
31244..31393
/feature="PKD1"
/number=20
/feature="putative"
31394..32327
/feature="PKD1"
/number=20
/feature="PKD1"
32328..32471
/feature="PKD1"
/number=21
/feature="putative"
32472..32598
/feature="PKD1"
/number=21
/feature="PKD1"
32599..32878
/feature="PKD1"
/number=21

```

```

/misc_feature
32616..33079
/notes="CpG island; putative"
/number=22
/feature="PKD1"
32879..32971
/feature="PKD1"
/number=22
/feature="putative"
32972..33185
/feature="PKD1"
/number=23
/feature="putative"
33186..33250
/feature="PKD1"
/number=23
/feature="putative"
33251..33410
/feature="PKD1"
/number=24
/feature="putative"
33411..33847
/feature="PKD1"
/number=24
/feature="putative"
33848..33952
/feature="PKD1"
/number=25
/feature="putative"
33953..36953
/feature="PKD1"
/number=25
/feature="putative"
36954..37098
/feature="PKD1"
/number=26
/feature="putative"
37007..37251
/notes="CpG island; putative"
/number=26
/feature="PKD1"
37099..37700
/feature="PKD1"
/number=26
/feature="putative"
37701..38330
/feature="PKD1"
/number=27
/feature="putative"
38331..38623
/feature="PKD1"
/number=27
/feature="putative"
38624..38780
/feature="PKD1"
/number=28
/feature="putative"
38781..38960
/feature="PKD1"
/number=28
/feature="putative"

```

```

exon      38961..39213
          /gene="PKD1"
          /map="16p13.3"
          /number=29
intron    39214..39337
          /gene="PKD1"
          /map="16p13.3"
          /number=29
exon      39338..39533
          /gene="PKD1"
          /map="16p13.3"
          /number=30
intron    39534..41027
          /gene="PKD1"
          /map="16p13.3"
          /number=30
repeat_region complement(39970..40228)
repeat_region /rpt_family="Alu"
exon      40311..40598
          /rpt_family="Alu"
          41028..41198
          /gene="PKD1"
          /map="16p13.3"
          /number=31
intron    41199..41284
          /gene="PKD1"
          /map="16p13.3"
          /number=31
exon      41285..41428
          /gene="PKD1"
          /map="16p13.3"
          /number=32
intron    41429..41522
          /gene="PKD1"
          /map="16p13.3"
          /number=32
misc_feature 41501..41730
          /note="CpG island; putative"
exon      41523..41733
          /gene="PKD1"
          /map="16p13.3"
          /number=33
intron    41734..41823
          /gene="PKD1"
          /map="16p13.3"
          /number=33
exon      41824..41950
          /gene="PKD1"
          /map="16p13.3"
          /number=34
intron    41951..43609
          /gene="PKD1"
          /map="16p13.3"
          /number=34
repeat_region 42463..42685
          /rpt_family="Alu"
repeat_region 43084..43321
          /rpt_family="Alu"
exon      43610..43726
          /gene="PKD1"
          /map="16p13.3"
          /number=35
intron    43727..43816
          /gene="PKD1"

```

```

          /map="16p13.3"
          /number=35
exon      43817..43866
          /gene="PKD1"
          /map="16p13.3"
          /number=36
intron    43867..44090
          /gene="PKD1"
          /map="16p13.3"
          /number=36
exon      44091..44275
          /gene="PKD1"
          /map="16p13.3"
          /number=37
intron    44276..44352
          /gene="PKD1"
          /map="16p13.3"
          /number=37
exon      44353..44446
          /gene="PKD1"
          /map="16p13.3"
          /number=38
intron    44447..47383
          /gene="PKD1"
          /map="16p13.3"
          /number=38
repeat_region 44778..45002
          /rpt_family="Alu"
repeat_region complement(46331..46602)
          /rpt_family="Alu"
exon      47384..47502
          /gene="PKD1"
          /map="16p13.3"
          /number=39
intron    47503..47580
          /gene="PKD1"
          /map="16p13.3"
          /number=39
exon      47581..47783
          /gene="PKD1"
          /map="16p13.3"
          /number=40
intron    47784..47855
          /gene="PKD1"
          /map="16p13.3"
          /number=40
exon      47856..48050
          /gene="PKD1"
          /map="16p13.3"
          /number=41
intron    48051..48500
          /gene="PKD1"
          /map="16p13.3"
          /number=41
exon      48501..48640
          /gene="PKD1"
          /map="16p13.3"
          /number=42
intron    48641..49001
          /gene="PKD1"
          /map="16p13.3"
          /number=42
exon      49002..49114
          /gene="PKD1"

```

intron	/map="16p13.3" /number=43 49115..49405 /gene="PKD1" /map="16p13.3" /number=43 49115..49405 /gene="PKD1" /map="16p13.3" /number=43 49406..49547 /gene="PKD1" /map="16p13.3" /number=43 49548..49687 /gene="PKD1" /map="16p13.3" /number=44 49688..49813 /gene="PKD1" /map="16p13.3" /number=45 49814..49996 /gene="PKD1" /map="16p13.3" /number=45 49997..50171 /gene="PKD1" /map="16p13.3" /number=46 50172..50418 /gene="PKD1" /map="16p13.3" /number=46 50419..50709 /gene="PKD1" /map="16p13.3" /number=47 50652..50653 /gene="PKD1" /map="16p13.3" /number=47 50710..50784 /gene="PKD1" /map="16p13.3" /number=47 50785..50919 /gene="PKD1" /map="16p13.3" /number=48 50796..50797 /gene="PKD1" /map="16p13.3" /number=48 51003..51308 /gene="PKD1"
misc_feature	49898..50854 /notes="CpG island; putative"
exon	49997..50171 /gene="PKD1"
intron	50172..50418 /gene="PKD1"
exon	50419..50709 /gene="PKD1"
conflict	50652..50653 /gene="PKD1" /notes="conflict with the seq /citation={1} 50710..50784 /gene="PKD1"
intron	50710..50784 /gene="PKD1"
exon	50785..50919 /gene="PKD1"
conflict	50796..50797 /gene="PKD1" /map="16p13.3" /notes="conflict with the seq /citation={1} 50920..51002 /gene="PKD1"
intron	50920..51002 /gene="PKD1"
exon	51003..51308 /gene="PKD1"

	intron	/map="16p13.3" number=49 51309..51398 /gene="PKO1" /map="16p13.3" number=49 51399..52883 /gene="PKO1" /map="16p13.3" number=50 51827..51828 /gene="PKO1" /map="16p13.3" note="conflict with the sequence in entry L33243" /citation=[1] 52720 conflict polyA_signal 52862..52867 /gene="PKO1" /map="16p13.3"
	exon	51399..52883 /gene="PKO1" /map="16p13.3" number=50 51827..51828 /gene="PKO1" /map="16p13.3" note="conflict with the sequence in entry L33243" /citation=[1] 52720 conflict polyA_signal 52862..52867 /gene="PKO1" /map="16p13.3"
	conflict	51827..51828 /gene="PKO1" /map="16p13.3" note="conflict with the sequence in entry L33243" /citation=[1] 52720 conflict polyA_signal 52862..52867 /gene="PKO1" /map="16p13.3"
	BASE COUNT	8486 a 17664 c 15766 g 11606 t
	ORIGIN	
DB 50;	Score	23; Match 63.9%; QryMatch 1.5%; Pred. No. 7.59e-01;
Matches	53; Conservative	0; Mismatches 30; Indels 0; Gaps 0;
Db	34485	cctccctcctcccctccccctccccctcctcttctccctccccctcccccc 34544
Qy	1090	CCTTCCTCCTTACCTGCCTTGCCGACCACCATGGATCTCACCCAGCTCCACCCC 1149
Db	34545	tctccctcctcctccccctcc 34567
Qy	1150	TGTTTCTCGACCTTCACCAACC 1172
RESULT 14		
LOCUS	ADRHUMJ	329 bp DNA VRL 22-SEP-1986
DEFINITION	Adenovirus type 12/human recombinant viral DNA, recombination junction.	
ACCESSION	M12410	
KEYWORDS	Recombinant adenovirus type 12 viral DNA, clone 4al.1.	
SOURCE	Recombinant adenovirus h12	
ORGANISM	Mastadenovirus h12	
REFERENCE	1 (bases 1 to 329)	
AUTHORS	Deuring,R. and Doerfler,W.	
TITLE	Proof of recombination between viral and cellular genomes in human KB cells productively infected by adenovirus type 12: Structure of the junction site in a symmetric recombinant (STREC)	
JOURNAL	Gene 26, 283-289 (1983)	
MEDLINE	84159511	
COMMENT	Clean copy sequence for [1] kindly provided by R.Deuring and W.Doerfler, 15-JUL-1986.	
FEATURES	NCBI gi: 209941	
source	Location/Qualifiers	
	1..329	
BASE COUNT	56 a 98 c 131 g 44 t	
ORIGIN	Real site.	
DB 67;	Score	23; Match 68.3%; QryMatch 1.5%; Pred. No. 7.59e-01;
Matches	43; Conservative	0; Mismatches 20; Indels 0; Gaps 0;

94 cggtcaggacaaaggcgagtgggccagggggctgggccctggaactggagggagtgga 153
 400 CGGAACAGACGCAAGGCACAGCACATTCATGGGGCAGCTGCCCTTCTACTGGAGGGAGTGA 459

Db 154 agg 156
Ov 460 TGG 462

RESULT	15		
LOCUS	CHKOVAL5A	5322 bp	DNA
DEFINITION	Chicken ovalbumin gene, 5' flank.		
			VRT
			15-MAR-1990

ACCESSION M29020
KEYWORDS ovalbumin.
SOURCE Chicken DNA, clones pBR EcoRI(6,7)
ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 5322)

AUTHORS Kaye, J.S.
JOURNAL Unpublished (1989) U. Rochester, Dept. Biol., Rochester, NY 14527
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted by J.S.Kaye, 14-OCT-1989.

```

NCBI gi: 212507
FEATURES
    source
        Location/Qualifiers
            1..5322
                /organism="Gallus gallus"

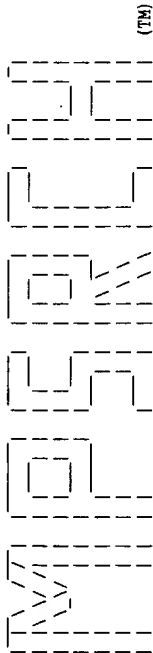
```

BASE COUNT	1469 a	1240 c	1103 g	1510 t
ORIGIN				

DB 73; Score	23; Match 87.1%; QryMatch 1.5%; Pred. No. 7.59e-01;
Matches	27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 3442 gtactgcttgcacagctgtgctgggcagg 3472
||||||| | ||||| | |||||
Cp 730 GTAAGCTTGGGACAGCTGTGTTGGCAGG 700

Search completed: Thu Sep 28 00:03:27 1995
Job time : 844 secs.



Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPPerch nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 27 23:38:11 1995; MasPar time 274.75 Seconds
824.986 Million cell updates/sec
Tabular output not generated.

```

Title:
Description:
Perfect Score:
N. A. Sequence:
Comp:

```

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 225244 seqs, 76266140 bases x 2

EST

1	EST1
2	EST2
3	EST3
4	EST4
5	EST5
6	EST6
7	EST7
8	EST8
9	EST9
10	EST10
11	EST11
12	EST12
13	EST13
14	EST14
15	EST15
16	EST16
17	EST17
18	EST18
19	EST19
20	EST20
21	EST21
22	EST22
23	EST23
24	EST24
25	EST25
26	EST26
27	EST27

Statistics: Mean 11.414; Variance 2.048; scale 5.574

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				DB	ID	Description	Pred. No.
	No.	Score	Match Length					
1	25	1.7	398	37	T16981	NIB2099-5R Homo sapie	1.10e-07	
2	23	1.5	436	19	R17453	vg14d02.r1 Homo sapie	4.44e-05	
3	22	1.5	368	8	HSCI1WE031	H. sapiens partial cd	7.53e-04	
c	4	22	1.5	411	R00328	ye96b08.r1 Homo sapie	7.53e-04	
5	22	1.5	236	24	R32925	vg62d09.r1 Homo sapie	7.53e-04	
6	22	1.5	361	32	R1CC0462A	Rice cDNA, partial se	7.53e-04	
c	7	22	1.5	405	R01838	ye84n03.r1 Homo sapie	7.53e-04	
8	22	1.5	355	18	R13668	vg60n02.r1 Homo sapie	7.53e-04	
c	9	22	1.5	352	T07692	ESF075582 Homo sapiens	7.53e-04	
10	22	1.5	198	37	T181179	0785c3 Plasmodium fal	7.53e-04	
11	22	1.5	313	52	T74922	ye58b01.r1 Homo sapie	7.53e-04	
c	12	21	1.4	430	R18325	vf75g03.r1 Homo sapie	1.13e-02	
13	21	1.4	383	47	T58482	yb61e11.r1 Homo sapie	1.13e-02	
c	14	21	1.4	346	16	R07221	yfi14b06.r1 Homo sapie	1.13e-02
c	15	21	1.4	413	R18134	vf55f05.r1 Homo sapie	1.13e-02	
c	16	21	1.4	496	25	R38202	ye93n05.s1 Homo sapie	1.13e-02
c	17	21	1.4	477	49	T63491	yc22f10.s1 Homo sapie	1.13e-02
c	18	21	1.4	418	R17243	vg12d04.r1 Homo sapie	1.13e-02	
c	19	21	1.4	337	5	H5BAGC082	H. sapiens partial cd	1.13e-02
c	20	21	1.4	461	25	R37851	ye98b02.s1 Homo sapie	1.13e-02
c	21	21	1.4	346	16	R07091	yfi13f06.r1 Homo sapie	1.13e-02

	22	21	1.4	302	32	RIC0791A	Rice cDNA, partial se	1.13e-02	
c	23	21	1.4	568	21	R16830	yf34a12.s1	Homo sapie	1.13e-02
c	24	21	1.4	356	23	R31294	yf60q12.s1	Homo sapie	1.13e-02
c	25	21	1.4	269	31	R59677	yh02q01.r1	Homo sapie	1.13e-02
c	26	21	1.4	375	24	R35940	yf68e06.r1	Homo sapie	1.13e-02
c	27	21	1.4	468	32	R63283	yf07f04.r1	Homo sapie	1.13e-02
c	28	21	1.4	472	27	R46605	yg51h10.s1	Homo sapie	1.13e-02
c	29	21	1.4	328	23	R32306	yf68b02.s1	Homo sapie	1.13e-02
c	30	21	1.4	352	52	T77909	yd16b02.r1	Homo sapie	1.13e-02
c	31	20	1.3	258	58	T97901	ye54b12.r1	Homo sapie	1.48e-01
c	32	20	1.3	312	58	T98862	ye45b10.s1	Homo sapie	1.48e-01
c	33	20	1.3	432	55	T87706	yf91h12.s1	Homo sapie	1.48e-01
c	34	20	1.3	278	9	HSC2HC102	H. sapiens partial cd	1.48e-01	
c	35	20	1.3	204	57	T93428	lmeST0196	Leishmania	1.48e-01
c	36	20	1.3	411	39	T27238	HCOEST5758	Homo sapiens	1.48e-01
c	37	20	1.3	436	16	R05578	yf87q06.r1	Homo sapie	1.48e-01
c	38	20	1.3	457	25	R38972	yd07d08.s1	Homo sapie	1.48e-01
c	39	20	1.3	475	29	T52272	yf73e11.r1	Homo sapie	1.48e-01
c	40	20	1.3	485	33	R1CS12189A	Rice cDNA, partial se	1.48e-01	
c	41	20	1.3	355	2	CEIK01607R	C.elegans cDNA clone	1.48e-01	
c	42	20	1.3	390	54	T83445	yd46c07.r1	Homo sapie	1.48e-01
c	43	20	1.3	539	53	T79953	yd85c11.s1	Homo sapie	1.48e-01
c	44	20	1.3	411	33	R1CR3020A	Rice cDNA, partial se	1.48e-01	
c	45	20	1.3	202	41	T33080	EST55609	Homo sapiens	1.48e-01

ALIGNMENTS

RESULT	1				
LOCUS		398 bp	mRNA	EST	01-AUG-1994
DEFINITION		NTB2099-5R Homo sapiens cDNA 5' end.			
ACCESSION		T16981			
KEYWORDS		EST.			
SOURCE		human library=Normalized infant brain, Bento Soares vector=BA, M13-derived host=E. coli DH5-alpha primer=M13 Reverse Reiter-HindIII Rsite2=NotI The normalized infant brain library, constructed by Bento Soares, Columbia University, was oligo-(dT) primed and directionally cloned into an M13-derived plasmid using total brain mRNA from a 72-day old human female afflicted with spinal muscular atrophy. The library was normalized as described elsewhere.			

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Stevens, T.J., Berry, R., Goold, R., Walter, N.A.R., Wilcox, A.S.,
Hopkins, J.A.A., Rubano, T., Weber, J., Soares, M.B. and Sikeia, J.M.
TITLE Gene-based STSS as the basis for a human gene map
JOURNAL Unpublished (1994)

Contact: Sikela JM
Department of Pharmacology
University of Colorado Health Sciences Center
Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
Tel: 3032706884
Fax: 3032707097
Email: tjs@tally.hsc.colorado.edu.

```

NCBI gi: 519143
FEATURES
            source
            Location/Qualifiers
            1..398
            /organism="Homo sapiens"
            /note="human"

```

BASE COUNT	100 a	117 c	75 g	101 t	5 others
------------	-------	-------	------	-------	----------

ORIGIN							
DB 37;	Score	25;	Match 62.1%;	QryMatch 1.7%;	Pred. No. 1.10e-07;		
Matches	64;	Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;	
Db	70	ctctctctctgtgagttgaccagcagcagcagctgtgcgaccaccagcaccatctctctctc	129				
Qy	1095	CTCTCTTTCACCTGCTTGTCCGACACCCATCGATCTCCACCCGCTCGACCCGCTGTTT	1154				
Db	130	ccagcagctctccagaaccatgccaggtctctctgctccatc	172				
Qy	1155	CTGACCCITTCGACACCATGCTTAAGTCTACCGCCGCTCATC	1197				
RESULT	2						
LOCUS	R17453	436 bp	mRNA	EST			
DEFINITION	Yg1402.r1 Homo sapiens cDNA clone 32427 5'.						14-APR-1995
ACCESSION	R17453						
KEYWORDS	EST.						

SOURCE

human clones-32427 library=Soares infant brain IN1B vector=Lafmid BA host=DH10B (ampicillin resistant) primer=M13RP1 Reitel=Not I Rite2=Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was prepared with a Not I - oligo(dT) primer [5' AACTGGAGAATTCGCGCCGACGAGATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 436)

REFERENCE

AUTHORS

Holman, H., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Hollman, L., Luitman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and
Wilson, R.

TITLE	JOURNAL	COMMENT
The WashU-Merck EST Project	Unpublished (1995)	

GDB: G00-404-774
Contact: Willson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 399
Source: IMAGE Consortium, LIML
This clone is available royalty-free through LIML; contact
IMAGE Consortium (info@image.liml.gov) for further information.

```

NCBI gi: 771063
FEATURES
    source
        Location/Qualifiers
            1..436
                /organism="Homo sapiens"
                /clone="32427"
                /note="human"

```

BASE COUNT	84 a	128 c	121 g	100 t	3 others
ORIGIN					

DB 19; Score 23; Match 69.5%; QryMatch 1.5%; Pred. No. 4.44e-05;

-	Matches	41;	Conservative	0;	Mismatches	18;	Indels	0;	Gaps	0;
---	---------	-----	--------------	----	------------	-----	--------	----	------	----

Dn 246 tcctgtgcctgctccccaccacccagcaggcgccctcccattccccctccctctgtttt 304
||||| ||||| ||||| | | | | ||||| || ||| |||||

- Qy 1097 TCCTTACAGTCGCTTGCCCCACCACCATGCATGCCACCCAGCTCACCCCCTGTTC 1155

```

RESULT      3
LOCUS       HSC1MEU031          368 bp    RNA            EST           06-NOV-1994
DEFINITION   H. sapiens partial cDNA sequence; clone c-lwe03.
ACCESSION   Z44264
KEYWORDS     partial cDNA sequence; transcribed sequence fragment.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; mitochondrial eukaryotes; Metazoa/Eumycota group;
             Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
             Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
             Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
             Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 368)
AUTHORS     Genexpress.
TITLE        Direct Submission
JOURNAL      Submitted (24-OCT-1994) to the EMBL/GenBank/DDJB databases.
GENETON     Genethon, B.P. 60, 91002 Evry Cedex France and Genetique
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801
Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE   2 (bases 1 to 368)
AUTHORS     Genexpress.
TITLE        The Genexpress cDNA program
JOURNAL      Unpublished
REFERENCE   3 (bases 1 to 368)
AUTHORS     Autray,C., Behar,G., Bois,F., Boucher,C., da Silva,C.,
Deignies,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Marige-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE       Integrated molecular analysis of the human genome and its
expression
C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
Clone library from B.Souares, Psychiatry Dept. Columbia University
USA;

Cloning method: total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
vector;
Sequencing method: single read, full automatic;
Primer: M13_reverse
cDNA sequence colinear to mRNA
Stretch removed: nothing
Normalization method: Bento Soares, P.N.A.S in press;
Genexpress_library_id: C;
Genexpress_sequence_id: ylc-lwe03;

No significant homology found with :
genbank release 81 swissprot release 28.
```

```
NCBI gi: 573389
Location/Qualifiers
1..368
 /organism="Homo sapiens"
 /dev_stage="3 months old"
 /isolate="muscular atrophy patient "
 /tissue_type="total brain"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
```

```
FEATURES
source
```

BASE COUNT	64 a	97 c	101 g	103 t	3 others
ORIGIN					
DB 8; Score	22; Match 86.7%; QryMatch 1.5%;	Pred. No. 7.53e-04;			
Matches	26; Conservative 0; Mismatches 4;	Indels 0; Gaps 0;			
Db 141	tctctctccagcccccactccctgttccct 170				
QY 1128	TCTCCACCCAGCTCCACCCCGCTGTTCCT 1157				
RESULT 4					
LOCUS	R00328	411 bp	mRNA	EST	31-MAR-1995
DEFINITION	ye9b08.r1 Homo sapiens cDNA clone 122967 5' similar to gb:W83653				
ACCESSION	RED CELL ACID PHOSPHATASE 1, ISOZYME F (HUMAN);				
KEYWORDS	R00328				
SOURCE	EST.				
human clone=122967 library=Soares fetal liver spleen lNFLS					
vector=pf7T3D (Pharmacia) with a modified polylinker host=DH10B					
(ampicillin resistant) primer=M13R1 Reitel=Pac I Rsite2=Eco RI					
Liver and spleen from a 20 week-post conception male fetus. 1st					
strand cDNA was primed with a Pac I - oligo(dT) primer [5'					
AAGCTGACAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded					
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac					
I and cloned into the Pac I and Eco RI sites of the modified pT733					
vector. Library went through one round of normalization. Library					
constructed by Bento Soares and M.Fatima Bonaldo.					
Homo sapiens					
ORGANISM					
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;					
Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 411)					
REFERENCE					
AUTHORS					
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,					
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,					
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,					
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and					
Wilson,R.					
TITLE					
JOURNAL					
COMMENT					
The WashU-Merck EST Project					
Unpublished (1995)					
Contact: Wilson RK					
WashU-Merck EST Project					
Washington University School of Medicine					
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108					
Tel: 314 286 1800					
Fax: 314 286 1810					
Email: est@watson.wustl.edu					
High quality sequence stops: 228					
Source: IMAGE Consortium, LNL					
This clone is available royalty-free through LNL; contact the					
IMAGE Consortium (info@image.lnl.gov) for further information.					
NCBI gi: 750064					
FEATURES					
source					
Location/Qualifiers					
1..411					
/organism="Homo sapiens"					
/clone="122967"					
/note="human"					
BASE COUNT	106 a	84 c	92 g	125 t	4 others
ORIGIN					
DB 14; Score	22; Match 80.6%; QryMatch 1.5%;	Pred. No. 7.53e-04;			
Matches	29; Conservative 0; Mismatches 7;	Indels 0; Gaps 0;			
Db 179	tcccagtgctacctctctctctccaggccctttt 214				

```

- Cp 1384  TCTCCAGTGTGTGTATCCCTTCCCGAGGGCCTTT 1349
      II IIIIIIIII III IIIIIII IIIIIIIIIII
- RESULT 5
LOCUS      R35295      236 bp      mRNA      EST      02-MAY-1995
DEFINITION Yg62d09.r1 Homo sapiens cDNA clone 37553 5'.
ACCESSION R35295
KEYWORDS   EST.
SOURCE     human clone-37553 library=Soares infant brain INIB vector=Lafmid BA
host=DH10B (ampicillin resistant) primer=M13p1 Rsite=Not I
Rs1e2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTCGAAGATTCGGCGCCGAGCATTTTTTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.
ORGANISM   Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 236)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Treviski,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT
GDB: G00-410-094
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 180
Source: IMAGE Consortium, L1NL
This clone is available royalty-free through L1NL ; contact the
IMAGE Consortium (info@image.l1nl.gov) for further information.
NCBI gi: 792196
FEATURES
source      Location/Qualifiers
            1..236
            /organism="Homo sapiens"
            /clone="37553"
            /note="human"
BASE COUNT  38 a 75 c 68 g 47 t 8 others
ORIGIN
DB 24; Score 22; Match 66.7%; QryMatch 1.5%; Pred. No. 7.53e-04;
Matches 38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Db 95 tectgtcctncttncaccacagcagtcgccccctccccctccccctctgtt 151
      IIII III IIIIIIIII III IIIIIII IIIIIIIIIII
Qy 1097 TCTTTCACCTGCCTTGCCACACCCATCGATTCACCCCGAGCTCCACCCCTGTT 1153
- RESULT 6
LOCUS      RIC0462A 361 bp      mRNA      EST      25-MAY-1995
DEFINITION Rice cDNA, partial sequence (C0462 1A).

```

<hr/>					
ACCESSION	D22591				
KEYWORDS	EST(expressed sequence tag).				
SOURCE	Oryza sativa (strain Nipponbare,)	Callus Callus cDNA to mRNA.			
ORGANISM	Oryza sativa				
Eukaryotes; mitochondrial eukaryotes; Chlorophyta/Embryophyta group; Charophyta/Embryophyta group; Embryophyta; Magnoliophyta; Liliopsida; Commelinidae; Poales; Poaceae; Oryza.					
REFERENCE	1 (bases 1 to 361)				
AUTHORS	Minobe, Y. and Sasaki, T.				
TITLE	Rice cDNA from callus				
JOURNAL	Unpublished (1993)				
COMMENT	Submitted (2-NOV-1993) to DDBJ by:				
	Yuzo Minobe				
	Dept. Rice Genome Research Program				
	National Institute of Agrobiological Resources				
	Kannondai 2-1-2				
	Tsukuba, Ibaraki				
	Japan				
	Phone: 0298-38-7441				
	Fax: 0298-38-7468				
	PROJECT = 'RGP' .				
<hr/>					
FEATURES	NCBI gi: 425911	Location/Qualifiers			
source	1..361	/organism='Oryza sativa'			
	/strain='Nipponbare'	/dev stage='Callus'			
	/sequenced_mol='cDNA to mRNA'	/tissue type='Callus'			
BASE COUNT	65 a 137 c 101 g 42 t 16 others				
<hr/>					
ORIGIN					
DB 32;	Score 22; Match 62.8%; QryMatch 1.5%; Pred. No. 7.53e-04;				
Matches 59;	Conservative 0; Mismatches 34; Indels 1; Gaps 1;				
<hr/>					
Db 50	gatgaacacacacctctgctccactcacctaccgtcgctctanagccgtacnccacaccca 109				
Qy 1077	GATGGACACACACCTTCCTCCTTCCACTGCGTTGCCACCACCATGGATCTCCACC- 1135				
<hr/>					
Db 110	ccccgcgcacctccgctccatcgccgccacacc 143				
Qy 1136	CAGCTCACCCCCCTGTTCTTGACCTTCGCC 1169				
<hr/>					
RESULT 7					
LOCUS	R01838	405 bp	mRNA	EST	31-MAR-1995
DEFINITION	ye84h03.r1 Homo sapiens cDNA clone 124469 5' similar to gb:M83653				
	RED CELL ACID PHOSPHATASE 1, ISOZYME F (HUMAN); .				
ACCESSION	R01838				
KEYWORDS	EST.				
SOURCE	human clone=124469 library=Soares fetal liver spleen INFLS vector=pT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13RP1 Rsite=Pac I Rsite=Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGAGAATTAAATAAGATCTTTTTTTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.				
ORGANISM	Homo sapiens				
	Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				

DB 36; Score 22; Match 78.4%; QryMatch 1.5%; Pred. No. 7.53e-04; Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

- Db 80 ggtgaggaagntcaggagccaggagttagatgg 116
 ||||| ||||| ||||| || ||||| |||||

Cp 1172 GCTGCTGACAGGCTCAGGAACAGGGGTGACCTGG 1136

RESULT 10

LOCUS T18179 198 bp mRNA EST 30-AUG-1994

DEFINITION 0785c3 Plasmodium falciparum cDNA clone 0785c 5'.

ACCESSION T18179

KEYWORDS EST.

SOURCE malaria parasite clones-0785c library=czapFDd2.1, Debopam Chakrabarti strain=Dd2 vector=Lambda ZAP II host=E. coli XL-1 blue primer=T3 Rsite=EcoR I Rsite2=Xho I PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dT-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector.

ORGANISM Plasmodium falciparum

Eucaryotae; Protozoa; Apicomplexa; Sporozoa; Coccidia; Eucoccidiida; Haemosporina; Plasmodium.

REFERENCE 1 (bases 1 to 198)

AUTHORS Chakrabarti, D., Reddy, G.R., Dame, J.B., Almira, E.C., Laipis, P.J., Ferl, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.

TITLE Analysis of Expressed Sequence Tags from Plasmodium falciparum

JOURNAL Mol. Biochem. Parasitol. 66, 97-104 (1994)

MEDLINE 95075403

COMMENT Contact: Debopam Chakrabarti
 Biotechnology & Infectious Diseases
 University of Florida
 Box 110880, Gainesville, FL 32611
 Tel: 9043924700 ext. 5817
 Fax: 9043929704
 Email: dchak@icbr.ifas.ufl.edu.

NCBI gi: 462965

FEATURES

source Location/Qualifiers

1..198

/organism="Plasmodium falciparum"

/clone="0785c"

/strain="Dd2"

/note="malaria parasite"

BASE COUNT 70 a 15 c 28 g 84 t 1 others

ORIGIN

DB 37; Score 22; Match 73.9%; QryMatch 1.5%; Pred. No. 7.53e-04; Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 92 atggttaacattatgaagctatttatttaccctttaagtatat 137
 || ||||| ||||| || ||||| ||||| ||||| ||||| |||||

Qy 1384 ATTGTAATAATTTTAGGAGCTATTTTTTTTAACCTATCAGCAATAT 1429

RESULT 11

LOCUS T74922 313 bp mRNA EST 02-MAR-1995

DEFINITION yc58b01.r1 Homo sapiens cDNA clone 84841 5'.

ACCESSION T74922

KEYWORDS EST.

SOURCE human clone-84841 library=Stratagene liver (#937224)
 vector=pBluescript SK host=SOIR cells (kanamycin resistant)
 primer=M13Rp1 Rsite=EcoRI Rsite2=XhoI Cloned unidirectionally.
 Primer: Oligo dT. Hepatectomy from normal 49 year old male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTGGCAGCAG-3'; 3' adaptor sequence: 5'-CTCCAGTTTTTTTTTTTTTTT-3'.

ORGANISM Homo sapiens

Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 313)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 253

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 691597

FEATURES

source Location/Qualifiers

1..313

/organism="Homo sapiens"

/clone="84841"

/note="human"

BASE COUNT 60 a 76 c 83 g 92 t 2 others

ORIGIN

DB 52; Score 22; Match 86.7%; QryMatch 1.5%; Pred. No. 7.53e-04; Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 110 tctctccccagcccaactccctgttcct 139
 ||||| ||||| ||||| ||||| ||||| |||||

Qy 1128 TCTCCAGCCCGAGCTCCAGCCCGCTGTTTCCT 1157

RESULT 12

LOCUS R13325 430 bp mRNA EST 12-APR-1995

DEFINITION YF75g03.r1 Homo sapiens cDNA clone 28121 5'.

ACCESSION R13325

KEYWORDS EST.

SOURCE human clone-28121 library=Soares infant brain INIB vector=Lafmid BA host=Dh10B (ampicillin resistant) primer=M13Rp1 Rsite=Not I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGAAGAAATTCGGCGCGCAGGAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens

Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

REFERENCE
AUTHORS
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 430)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
GDB: G00-400-468
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 324
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE
JOURNAL
COMMENT

FEATURES
source
NCBI gi: 766401
Location/Qualifiers
1..430
/organism="Homo sapiens"
/clone="28121"
/note="human"

BASE COUNT
ORIGIN
91 a 113 c 121 g 100 t 5 others
DB 18; Score 21; Match 59.1%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Db 212 aagaggccctaaactcccccatccagatggatggagccctccaaaggacct 266
||||| ||||| | | ||||| ||||| || ||||| | |||||
Cp 1406 AATAGCTCTAAATTTTACAAATCTCCAGTGTCTGTATCCCTCCCGAGGCCT 1352

RESULT 13
LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
human clone=75692 library=Stratagene ovary (#937217)
vector=Bluescript SK host=SOLR cells (kanamycin resistant)
primer=W13RPI Rsite1=EcoRI Rsite2=XhoI Cloned unidirectionally.
Primer: Oligo dt. Total ovary tissue, normal 49 year old caucasian female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTGGCAGCAG-3'; 3' adaptor sequence: 5'-CTCGAGCTTTTCTTTTCTTTT-3'.
Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 383)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
WashU-Merck EST Project
Unpublished (1995)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 197
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 660319
Location/Qualifiers
1..383
/organism="Homo sapiens"
/clone="75692"
/note="human"

BASE COUNT
ORIGIN
112 a 86 c 92 g 88 t 5 others
DB 47; Score 21; Match 75.7%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Db 283 caggngcagccctttccaggagctnctttagtta 319
||||| ||||| | ||||| ||||| |||||
Qy 123 CAGGGGAGAGGCCCCATACAGGGAGCCACTTCAGTTA 159

RESULT 14
LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
human clone=126803 library=Soares fetal liver spleen INFLS
vector=PTT3D (Pharmacia) with a modified polylinker host=DH10B
(ampicillin resistant) primer=W13RPI Rsite1=Fac I Rsite2=Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo(dT) primer [5'
AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.
Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 211

Source: IMAGE Consortium, L1NL
This clone is available royalty-free through L1NL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source
1..346
/organism="Homo sapiens"
/clone="126803"
/note="human"

BASE COUNT 95 a 63 c 110 g 73 t 5 others

ORIGIN

DB 16; Score 21; Match 88.9%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 202 tcagaagtccaggccaagctgtgtcgg 228

|||||

Cp 831 TCAGAAGTCCAGGCGCAGCAGTTCTGG 805

RESULT 15

LOCUS R12342 413 bp mRNA EST 12-APR-1995
DEFINITION yf55f05.r1 Homo sapiens cDNA clone 25883 5' similar to gb:M10058
ASTA LOGLYCOPROTEIN RECEPTOR 1 (HUMAN);

ACCESSION R12342

KEYWORDS EST.

SOURCE human clone=25883 library=Soares infant brain lNIB vector=Lafmid BA
host=DH10B (ampicillin resistant) primer=M13Rp1 Rsite1=Not I
Rs1te2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo (dT) primer [5'
AAGTCGACGAATTCGGCGCGCAGGAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM

Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 413)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE

JOURNAL

COMMENT

GDB: G00-398-230
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 329

Source: IMAGE Consortium, L1NL

This clone is available royalty-free through L1NL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 765418

Location/Qualifiers

FEATURES

source 1..413

/organism="Homo sapiens"

/clone="25883"

/note="human"

BASE COUNT 73 a 137 c 109 g 92 t 2 others

ORIGIN

DB 18; Score 21; Match 69.2%; QryMatch 1.4%; Pred. No. 1.13e-02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 90 aggttcnctctgtcccgcgctctctgcacccccctctctctgtgagggccac 141

|||||

Cp 486 AGGGTTCCAACCTGTCCTGCTGCGCATCATCTCCCTCCAGTAGAAGGGACAC 435

Search completed: Wed Sep 27 23:43:35 1995

Job time : 324 secs.

MAPSREL (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 27 11:41:12 1995; MasPar time 5.69 Seconds

Tabular output not generated. 423.266 Million cell updates/sec

Title: >US-08-252-491-2
Description: (1:379) from US08252491.pep
Perfect Score: 2644
Sequence: 1 MAPCK10GRPTQGATSVRH.....STAPHPVTMYPHRNLSQET 379

Scoring table: PAM 150
Gap 11

Searched: 53402 seqs, 6354270 residues

Database: a-geneseq
1 a-gen1
2 a-gen2
3 a-gen3
4 a-gen4
5 a-gen5
6 a-gen6
7 a-gen7
8 a-gen8
9 a-gen9
10 a-gen10

Statistics: Mean 34.679; Variance 190.574; scale 0.182

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	5.0	193	3	P50300 Human erythropoietin	3.02e-01
2	131	5.0	330	4	R23076 Epo:IL-3 short, recom	3.02e-01
3	131	5.0	349	4	R23079 Epo:IL-3 Flex, recom	3.02e-01
4	131	5.0	193	2	R11892 Erythropoietin analog	3.02e-01
5	131	5.0	193	2	P70256 Sequence of human ery	3.02e-01
6	131	5.0	188	1	P81195 Erythropoietin encode	3.02e-01
7	131	5.0	193	3	P60597 Clone lambda HEPOFL13	3.02e-01
8	131	5.0	188	3	P60599 Clone lambda HEPOFL16	3.02e-01

9	131	5.0	193	2	R11858 Modified human erythr	3.02e-01
10	131	5.0	193	2	R11893 Erythropoietin analog	3.02e-01
11	131	5.0	193	2	R11891 Erythropoietin analog	3.02e-01
12	130	4.9	193	3	P50343 EPO encoded by clone	3.48e-01
13	130	4.9	193	3	P60598 Open reading frame co	3.48e-01
14	124	4.7	334	3	P83194 Sequence of a bioadhe	8.17e-01
15	118	4.5	193	2	R11859 Erythropoietin analog	1.90e+00
16	116	4.4	334	1	P82971 Bioadhesive precursor	2.50e+00
17	114	4.3	192	3	P50301 Monkey erythropoietin	3.30e+00
18	109	4.1	254	2	R10531 Prod. of DNA of pMG07	6.55e+00
19	108	4.1	372	9	R47260 Pre-pro-GDF-1.	7.51e+00
20	108	4.1	372	9	R50207 Human GDF-1(fx) morph	7.51e+00
21	108	4.1	372	9	R46750 Human morphogenic pro	7.51e+00
22	108	4.1	372	9	R60961 Human GDF-1(fx) morph	7.51e+00
23	108	4.1	372	9	R57980 Human GDF-1(fx) morph	7.51e+00
24	108	4.1	372	9	R47297 GDF-1.	7.51e+00
25	108	4.1	372	6	R34582 Human GDF-1 morphogen	7.51e+00
26	108	4.1	372	6	R33942 Morphogen GDF-1 full	7.51e+00
27	108	4.1	302	4	R23596 Recombinant hematopoi	7.51e+00
28	108	4.1	372	4	R22376 hGDF-1.	7.51e+00
29	108	4.1	121	5	R29164 PRP3.	7.51e+00
30	108	4.1	321	4	R23075 IL-3:Epo short, recom	7.51e+00
31	106	4.0	151	5	R29165 PRP3 (from genomic cl	9.84e+00
32	105	4.0	77	1	P81196 Erythropoietin encode	1.13e+01
33	104	3.9	156	2	R10532 Prod. of DNA of pMG08	1.29e+01
34	103	3.9	340	4	R23078 IL-3:Epo Flex, recom	1.47e+01
35	102	3.9	321	4	R23597 Recombinant hematopoi	1.47e+01
36	102	3.9	114	5	R29163 PRP2.	1.68e+01
37	101	3.8	533	7	R39468 hRXR-beta1.	1.92e+01
38	100	3.8	521	9	R57772 c-MEF2.	2.19e+01
39	99	3.7	322	4	R23599 Recombinant hematopoi	2.50e+01
40	99	3.7	167	3	P50298 Human recombinant ery	2.50e+01
41	99	3.7	166	2	P70398 Sequence of human ery	2.50e+01
42	99	3.7	167	3	P50299 Human recombinant ery	2.50e+01
43	99	3.7	166	4	R23593 Recombinant hematopoi	2.50e+01
44	99	3.7	303	4	R23598 Recombinant hematopoi	2.50e+01
45	97	3.7	439	5	R28150 Sugar beet chitinase	3.26e+01

ALIGNMENTS

RESULT	1
ID	P50300 standard; protein; 193 AA.
AC	P50300;
DT	01-JAN-1980 (first entry)
DE	Human erythropoietin encoded by positive clone (phage lambda-hEl)
DE	Isolated from human fetal liver gene bank.
KW	Erythropoietin; red blood cell; erythrocyte; anaemia; blood;
KW	disorder; ss; phage lambda-hEl; gene bank.
OS	Homo sapiens.
PN	W08502610-A.
PD	20-JUN-1985.
PF	11-DEC-1984; U02021.
PR	13-DEC-1983; US-561024.
PR	21-FEB-1984; US-582185.
PR	28-SEP-1984; US-655841.
PR	30-NOV-1984; US-675298.
PA	(KIRI-) KIRIN-AMGEN INC.
DR	N-PSDB; N50347.
PT	New polypeptide having properties of erythropoietin - is prepd.
PT	by cultivation of transformed eucaryotic or procaryotic host
PS	Disclosure; Page 43; 113pp; English.
CC	Human erythropoietin encoded by a sequence encoded by this phage

ID R11892 standard; Protein; 193 AA.
AC R11892;
DT 22-JUL-1991 (first entry)
DE Erythropoietin analogue [Thr1125].
KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
KW sialic acids; glycosylation.
OS Homo sapiens.
PN W09105867-A.
PN EP-428267-A.
PD 02-MAY-1991.
PF 09-OCT-1990; U05758.
PR 13-OCT-1989; US-421444.
PA (AMGE-) AMGEN INC.
PI Strickland TW, Byrne TE, Elliott SG;
DR WPI; 91-148745/20.
DR WPI; 91-150265/21.
PT Recombinant erythropoietin iso-forms and purification. - increase
PT haematocrit levels in mammals and contg. specific number of
PT sialic acids
PS Claim 33; Page 45; 60pp; English
CC The analogue was constructed to add an O-glycosylation site
CC at Thr125 (amino acid 154 in this sequence) and was produced by
CC expression of DNA obtained by site-directed mutagenesis of DNA
CC encoding EPO. The analogue has a higher sialic acid content than
CC human EPO and has increased biological activity.
CC It can be used to treat mammals to cause bone marrow cells to
CC increase prodn. of reticulocytes and red blood
CC thereby increasing haematocrit levels.
CC See also R11859, R11891-93.
SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 12 lllsllslplglpvgapprrlclcdsrlylleakeaenittgcaehcslnenitvptd 71
|||: : l::: |::| ||:|:|: |::: : |::: |: : |
Qy 29 LLLAAMLLAVARLTUSSVPAPACDPRLLKLLRDSLLHLSRLSQCPDVPDPLSPVLLPAV 88
:: || |::: ||:| |::: ||:| |::: ||
Db 72 knfyawkrmevgqgavevwgglallseavlrqall 108
:: || |::: ||:| |::: ||:| |::: ||
Qy 89 DFSLGEWKTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 5
ID P70256 standard; protein; 193 AA.
AC P70256;
DT 19-FEB-1991 (first entry)
DE Sequence of human erythropoietin (EPO).
KW Renal anaemia therapy; hormone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..27
FT /label=SIGNAL 28..193
FT Protein 81..97
FT Region
FT /note="Fragment that probe N70361 is based on"
PN EP-232034-A.
PD 12-AUG-1987.
PF 19-JAN-1987; 300399.
PR 23-JAN-1986; JP-012868.
PA (SUMO) SUMITOMO CHEM IND KK.
PA (SUMI-) SUMITOMO SEIYAKU KK.
PI Yanagi H, Ogawa I, Okamoto M, Hozumi T, Soga A, Yoshima T,
Tateuchi M.

DR WPI; 87-223006/32.
DR N-PSDB; N70360, N70361.
PT Human erythropoietin prodn. - by culturing human cells, esp.
PT Namalwa cells, transfected with DNA encoding human erythropoietin
PS Disclosure; Fig 1; 22pp; English.
CC A cDNA library was prepd. from the poly (A) RNA, which was isolated
CC from the erythropoietin-producing human hepatoma cell Hp-1. The cDNA
CC library was screened using the probes given in N70361 and N70362. A
CC plasmid (named as p58-A20) was isolated. The nucleotide sequence of
CC the cDNA obtained from this clone is shown in N70360.
SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 12 lllsllslplglpvgapprrlclcdsrlylleakeaenittgcaehcslnenitvptd 71
|||: : l::: |::| ||:|:|: |::: : |::: |: : |
Qy 29 LLLAAMLLAVARLTUSSVPAPACDPRLLKLLRDSLLHLSRLSQCPDVPDPLSPVLLPAV 88
:: || |::: ||:| |::: ||:| |::: ||
Db 72 knfyawkrmevgqgavevwgglallseavlrqall 108
:: || |::: ||:| |::: ||:| |::: ||
Qy 89 DFSLGEWKTQTEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 6
ID P81195 standard; protein; 188 AA.
AC P81195;
DT 20-NOV-1990 (first entry)
DE Erythropoietin encoded by EPO 140B.
KW EPO; erythropoietin; anaemia; renal failure.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label=leader sequence 23..188
FT Protein
FT /label=EPO
FN EP-267678-A.
PD 18-MAY-1988.
PF 15-SEP-1987; 308130.
PR 15-SEP-1986; US-907369.
PA (INTE-) Integrated genetics.
PI Beck AK, Withy RM, Zabrecky JR, Massiello NC;
DR WPI; 88-134531/20.
DR N-PSDB; N81554.
PT Recombinant human erythropoietin - produced by a transformed rodent
PT capable of producing N-linked and O-linked glycosylated human
PT erythropoietin.
PS Disclosure; p; English.
CC EPO 104B was one of four positive clones isolated from a cDNA
CC library prepd. from mRNA extracted from a human foetus of about 20
CC wk. gestation. The clone was identified using two probes, EPO1 and
CC EPO2 based on the published sequence of EPO (Nature (1985) Vol.313,
CC p.806). The sequence between nucleotides 63 and 724 has 100% homo-
CC logy with the published sequence. It encodes the 166 AAs of the
CC mature EPO protein and 22 AAs of the leader sequence. This clone
CC and a second, EPO 125, were used to construct a full length clone
CC which was expressed in rodent epithelial cells.
CC See also P81196.
SQ Sequence 188 AA;

DB 1; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;
Db 7 lllsllslplglpvgapprrlclcdsrlylleakeaenittgcaehcslnenitvptd 66

Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRSHLLSHLSRLSQCPDVPDPLSIPVLLPAV 88
 |||: : |::: |::| |::|:|:|: |:: : : |:: : |: : |: |

Db 67 kmfyawkrmevggqavewgglallseavlrqgall 103
 :: || |:: |:: |::|:|:|: |::|

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 7

ID P60597 standard; protein; 193 AA.
 AC P60597;
 DE 01-JAN-1980 (first entry)
 DE Clone lambda HEPOFL13 encoding human erythropoietin.
 KW Erythropoietin; lambda HEPOFL13; recombinant plasmid vector; anaemia;
 KW mammal cell culture; 3T3; CHO; Chinese hamster ovary; ss.
 OS Homo sapiens.
 PN W08603520-A.
 PD 19-JUN-1986.
 PF 03-DEC-1985; U02405.
 PR 04-DEC-1984; US-677813.
 PR 03-JAN-1985; US-688622.
 PR 22-JAN-1985; US-693258.
 PA (GENE-) GENETICS INST INC.
 PA (FRIT/) FRITSCHE E.
 PI Fritsch E, Hewick RM, Jacobs K;
 DR WPI; 86-169459/26.
 DR N-PSDB; N60513.
 PT Prodn. of human cDNA clone expressing erythropoietin - for mass
 prodn. of erythropoietin, useful for treating anaemia
 PS Disclosure; Page 7; 61pp; English.
 CC A recombinant plasmid vector expressing this clone is expressed in e.
 CC g 3T3 or CHO cell cultures. The produced erythropoietin is useful
 CC for treatment of anaemia, especially renal anaemia. The cloned gene
 CC expresses high levels of the protein and thus provides a means of
 CC mass production. See also N60514-21 and P60598-99.
 CC Sequence 193 AA;

DB 3; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 lllslslplglpvgaprrlcsrvlrylleakeaenittgcaehcslnenitvpdt 71
 |||: : |::: |::| |::|:|:|: |:: : : |:: : |: : |: |

Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRSHLLSHLSRLSQCPDVPDPLSIPVLLPAV 88

Db 72 kmfyawkrmevggqavewgglallseavlrqgall 108

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 8

ID P60599 standard; protein; 188 AA.
 AC P60599;
 DE 01-JAN-1980 (first entry)
 DE Clone lambda HEPOFL16 encoding human erythropoietin.
 KW Erythropoietin; lambda HEPOFL16; recombinant plasmid vector; anaemia;
 KW mammal cell culture; 3T3; CHO; Chinese hamster ovary; ss.
 OS Homo sapiens.
 PN W08603520-A.
 PD 19-JUN-1986.
 PF 03-DEC-1985; U02405.
 PR 04-DEC-1984; US-677813.
 PR 03-JAN-1985; US-688622.
 PR 22-JAN-1985; US-693258.
 PA (GENE-) GENETICS INST INC.

PA (FRIT/) FRITSCHE E.
 PI Fritsch E, Hewick RM, Jacobs K;
 DR WPI; 86-169459/26.
 DR N-PSDB; N60519.

PT Prodn. of human cDNA clone expressing erythropoietin - for mass
 prodn. of erythropoietin, useful for treating anaemia
 PS Disclosure; Page 20; 61pp; English.
 CC A recombinant plasmid vector expressing this clone is expressed in e.
 CC g 3T3 or CHO cell cultures. The produced erythropoietin is useful
 CC for treatment of anaemia, especially renal anaemia. The cloned gene
 CC expresses high levels of the protein and thus provides a means of
 CC mass production. See also N60513-18, N60520-21 and P60598.
 CC Sequence 188 AA;

DB 3; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 7 lllslslplglpvgaprrlcsrvlrylleakeaenittgcaehcslnenitvpdt 66
 |||: : |::: |::| |::|:|:|: |:: : : |:: : |: : |: |

Qy 29 LLLAAMLLAVARLTSSVPACDPRLINKLRSHLLSHLSRLSQCPDVPDPLSIPVLLPAV 88

Db 67 kmfyawkrmevggqavewgglallseavlrqgall 103

Qy 89 DFSLGEMKQTQEQSKAQDILGAVSLLEGVMAARGQL 125

RESULT 9

ID R11838 standard; protein; 193 AA.
 AC R11838;
 DT 24-JUL-1991 (first entry)
 DE Modified human erythropoietin.
 KW EPO; anaemia.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 28..193
 FT /label= Mature modified EPO
 FT Misc.difference 51
 FT /label= May be absent or any AA except Asn
 FT Misc.difference 65
 FT /label= May be absent or any AA except Asn
 FT Misc.difference 110
 FT /label= May be absent or any AA except Asn
 FT EP-427189-A.
 PD 15-MAY-1991.
 PF 06-NOV-1990; 121187.
 PR 07-NOV-1989; JP-289143.
 PA (SNOW) SNOW BRAND MILK PRODUCTS.
 PI Yamaguchi K, Ueda M;
 DR WPI; 91-141727/20.

DR New human erythropoietin(s) modified at positions 24, 38 and 83 -
 PT having improved biological activity for stimulating erythrocyte
 PT differentiation and treating anaemia.
 PS Disclosure; Page 8-9; 14pp; English.
 CC Modified EPO has a reduced number of sialic acid binding sites, and
 CC the galactose residue is not exposed. By stripping away these sugars
 CC from the glycoprotein, the affinity of the EPO to its receptor is
 CC greatly enhanced. Thus, the modified EPO shows a higher activity and
 CC is useful in the treatment of anaemia.
 CC Sequence 193 AA;

DB 2; Score 131; Match 23.0%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 23; Conservative 32; Mismatches 45; Indels 0; Gaps 0;

Db 12 lllslslplglpvgaprrlcsrvlrylleakeaenittgcaehcslnenitvpdt 71

Qy 29 LLLAAMLLAVARLTSSPVAPACDPRLLNKLRLDLSHLLHSRLSQCPDVPDPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqgavewqglalleavlrqgallvqs 111
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLLEGVMAARGQLPS 128

RESULT 10

ID R11891 standard; Protein; 193 AA.
 AC R11891;
 DE 22-JUL-1991 (first entry)
 KW Human erythropoietin analogue [Pro]124, [Thr]125.
 KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
 KW sialic acids; glycosylation.
 OS Homo sapiens.
 PN W09105867-A.
 PN EP-428267-A.
 PD 02-MAY-1991.
 PF 09-OCT-1990; U05758.
 PR 13-OCT-1989; US-421444.
 PA (AMGE-) AMGEN INC.
 PI Strickland TW, Byrne TE, Elliott SG;
 DR WPI; 91-148745/20.
 DR WPI; 91-150265/21.
 PT Recombinant erythropoietin iso-forms and purificn. - increase
 PT haematocrit levels in mammals and contg. specific number of
 PT sialic acids
 PS Claim 33; Page 45; 60pp; English
 CC The analogue was constructed to add an O-glycosylation site
 CC at Thr125 (amino acid 154 in this sequence) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLSLLPLGLPVLGAPPRICDRVLYLLEAKEANITTCGAECALNENITVPDT 71
 Qy 29 LLLAAMLLAVARLTSSPVAPACDPRLLNKLRLDLSHLLHSRLSQCPDVPDPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqgavewqglalleavlrqgall 108
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLLEGVMAARGQL 125

RESULT 11

ID R11891 standard; Protein; 193 AA.
 AC R11891;
 DE 22-JUL-1991 (first entry)
 KW Human erythropoietin analogue [Asn]125, [Ser]127.
 KW Human erythropoietin; EPO; isoform; analogue; haematocrit;
 KW sialic acids; glycosylation.
 OS Homo sapiens.
 PN W09105867-A.
 PN EP-428267-A.
 PD 02-MAY-1991.
 PF 09-OCT-1990; U05758.

PR 13-OCT-1989; US-421444.
 PA (AMGE-) AMGEN INC.
 PI Strickland TW, Byrne TE, Elliott SG;
 DR WPI; 91-148745/20.
 DR WPI; 91-150265/21.
 PT Recombinant erythropoietin iso-forms and purificn. - increase
 PT haematocrit levels in mammals and contg. specific number of
 PT sialic acids
 PS Claim 33; Page 45; 60pp; English
 CC The analogue was constructed to add an N-glycosylation site
 CC at Asn125 (amino acid 154 in this sequence) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 3.02e-01;
 Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLSLLPLGLPVLGAPPRICDRVLYLLEAKEANITTCGAECALNENITVPDT 71
 Qy 29 LLLAAMLLAVARLTSSPVAPACDPRLLNKLRLDLSHLLHSRLSQCPDVPDPLSIPVLLPAV 88
 Db 72 kmfyawkrmevgqgavewqglalleavlrqgall 108
 Qy 89 DFSLGEKMTQTEQSKAQDILGAVSLLLEGVMAARGQL 125

RESULT 12

ID P50343 standard; Protein; 193 AA.
 AC P50343;
 DT 10-MAR-1992 (first entry)
 DE EPO encoded by clone lambda-HEPOFL13.
 KW Erythropoietin.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= signal_sequence
 FT Protein 27..193
 FT /label= mature_EPO
 PN DK8406107-A.
 PD 16-AUG-1985.
 PF 27-DEC-1984; 281862.
 PR 27-DEC-1983; US-565627.
 PR 27-DEC-1983; US-566057.
 PR 04-DEC-1984; US-677813.
 PR 21-DEC-1984; NO-005186.
 PR 21-DEC-1984; ZA-010034.
 PR 22-JAN-1985; US-693258.
 PR 22-JAN-1985; US-688622.
 PA (GENE-) GENETICS INST.
 PI Kaufman RJ.
 DR WPI; 85-318061/51.
 DR N-PSDB; N50443.
 PT Vector system for introducing heterologous DNA into eukaryotic
 PT cells - comprises prod. gene and accessory DNA for enhanced
 PT expression of heterologous protein by the cells.
 PS Disclosure; Fig 13; 62pp; Danish.
 CC See also US4740461 88.04.26 (8819) (first major country equivalent).
 CC The sequence is encoded by clone lambda-HEPOFL13 which contains

89 D F S L G E W K I Q T E Q S K A Q D I L G A V S L L E G V M A A R G Q L 125

PT Recombinant erythropoietin iso-forms and purificn. - increase
PT haematocrit levels in mammals and contg. specific number of

PT Recombinant erythropoietin iso-forms and purificn. - increase
PT haematocrit levels in mammals and contg. specific number of

PT sialic acids

PS Claim 33; Page 45; 60pp; English
 CC The [Asn]69 analogue was constructed to add an N-glycosylation site
 CC at Asn69 (amino acid 96 in this sequences) and was produced by
 CC expression of DNA obtained by site-directed mutagenesis of DNA
 CC encoding EPO. The analogue has a higher sialic acid content than
 CC human EPO and has increased biological activity.
 CC It can be used to treat mammals to cause bone marrow cells to
 CC increase prodn. of reticulocytes and red blood
 CC thereby increasing haematocrit levels.
 CC See also R11859, R11891-93.
 SQ Sequence 193 AA;

DB 2; Score 118; Match 21.6%; QryMatch 4.5%; Pred. No. 1.90e+00;
 Matches 21; Conservative 33; Mismatches 43; Indels 0; Gaps 0;

Db 12 lllslslplglplgpprlldsrlylleakeanittgcaehcslnenitvpdt 71

|||: : l::: |::| ||:|::| : : l:: :|:

Qy 29 LLLAALLAVARLTSSVPAPACDPRLLNKLRLDSHLHSLRSQCDDVDPLSIPVLLPAV 88

Db 72 knfyawkmevggqavevwqglanlseavlrqgall 108

:: || |:: ::: |l:: ::: |

Qy 89 DFLGKWKTEQSKAQDILGAVSLLEGVMAARGQL 125

Search completed: Wed Sep 27 11:41:31 1995
 Job time : 19 secs.

MASSRFL

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 27 11:40:12 1995; MasPar time 14.27 Seconds
Tabular output not generated.
619.737 Million cell updates/sec

Title: >US-08-252-491-2
Description: (1:379) from US08252491.pep
Perfect Score: 2644
Sequence: 1 MAPGKIQGRGTQIGATSVRH.....STAPHPVTMYPHRNLSQET 379

Scoring table: PAM 150
Gap 11

Searched: 77573 seqs, 23340141 residues

Database: pir44

1 ANNO1
2 ANNO2
3 ANNO3
4 UNANNO1
5 UNANNO2
6 UNANNO3
7 UNANNO4
8 UNANNO5
9 UNANNO6
10 UNANNO7
11 UNREV1
12 UNREV2

Statistics: Mean 46.880; Variance 139.889; scale 0.335

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description	Pred. No.
1	2482	93.9	356	12	thrombopoietin - mou	0.00e+00
2	1748	66.1	353	12	c-MPI ligand - human	2.71e-221
3	983	37.2	286	9	megakaryocyte growth	1.27e-113
4	145	5.5	192	12	erythropoietin - rat	9.44e-04
5	141	5.3	192	4	erythropoietin precu	2.42e-03
6	131	5.0	193	2	erythropoietin precu	2.41e-02

ALIGNMENTS

RESULT ENTRY	1
TITLE	S45330 #type complete
ORGANISM	thrombopoietin - mouse
DATE	#formal name Mus musculus #common name house mouse 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change 10-Dec-1994
ACCESSIONS	S45330
REFERENCE	S45330
#authors	Lok, S.; Kaushansky, K.; Holly, R.D.; Kuijper, J.L.; Lofton-Day, C.E.; Oort, P.J.; Grant, F.J.; Heipel, M.D.; Burkhead, S.K.; Kramer, J.M.; Bell, L.A.; Sprecher, C.A.; Blumberg, H.; Johnson, R.; Prunkard, D.; Ching, A.F.T.; Macheves, S.G.; Bailey, M.C.; Forstrom, J.W.; Buddle, M.M.; Osborn, S.G.; Evans, S.J.; Sheppard, P.O.; Presnell, S.R.; O'Hara, P.J.; Hagen, F.S.; Roth, G.J.; Foster, D.C.
#journal	Nature (1994) 369:565-568
#title	Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo.
#accession	S45330
#status	preliminary
#residues	1-356 ##label LOK

```

SUMMARY
DB 12; Score 2482; Match 100.0%; QryMatch 93.9%; Pred. No. 0.00e+00;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 meltlllvaamlavartlaspvaccdrllnkllrdshlhrlerlscpcvdpplapv 60
Qy 24 MELTDLLLAAMLLAVARLTSSVPACDPRLLNKLGRDHLHSLRLSCQPDVPLSIPV 83

Db 61 llpavdfsfgewktqtegskaqdlilgavllleqymaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSIGEMKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSGQVRL 143

Db 121 llgalqllgtqlpqrtrttahkdpnaiflqlgllrgkvrlflllvegptlcvrtppt 180
Qy 144 LLGALQLLGTQLPQGRTRTTAKDPNALFLSLIQQLLRGKVRFLLLVEGPTLCVRRTPPT 203

Db 181 avpsstsqlltlnkfmrtsqletnfsartagpqlslrlqgrfvkitpqdlnqtsrs 240
Qy 204 AVPSSTSQLTLNKFMRTSGLTWNFSVARTAGPGLSLRLQGRFVKITPQQLNQTSRS 263

Db 241 pvqlsglntlrthgvpvngthglfagtslqtleasdispgafnkgslafnlqggllppspila 300
Qy 264 PVQLSGLNTLRTHGVPVNGTHGLFAGTSLQTLLEASDISPGAFNKGSLAFNLQGGLLPPSPILA 323

Db 301 pdgthtppspalptthgspqqlhplfdpdtmptnstaphvmtymphprnlsoet 356
Qy 324 PDGHTPPSPALPTTHGSPQQLHPLFDPDTMPTNSTAPHVMTYMPHPRNLSEOT 379

RESULT 2
ENTRY S45331 #type complete
TITLE c-Mpl ligand - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Dec-1994; #sequence_revision 10-Dec-1994; #text_change 10-Dec-1994

ACCESSIONS S45331
REFERENCE #authors de Sauvage, F.J.; Hass, P.E.; Spencer, S.D.; Malloy, B.E.; Gurney, A.L.; Spencer, S.A.; Darbonne, W.C.; Henzel, W.J.; Wong, S.C.; Kuang, W.J.; Oles, K.J.; Hultgren, B.; Solberg Jr., L.A.; Goeddel, D.V.; Eaton, D.L.
#journal Nature (1994) 369:533-538
#title Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand.
#accession S45331
#status preliminary
#residues 1-353 #label DEA
SUMMARY #length 353 #molecular-weight 37822 #checksum 1597

DB 12; Score 1748; Match 71.6%; QryMatch 66.1%; Pred. No. 2.71e-221;
Matches 255; Conservative 47; Mismatches 49; Indels 5; Gaps 4;

Db 1 meltlllvaamlavartlaspvaccdrllnkllrdshlhrlerlscpcvdpplapv 60
Qy 24 MELTDLLLAAMLLAVARLTSSVPACDPRLLNKLGRDHLHSLRLSCQPDVPLSIPV 83

Db 61 llpavdfsfgewktqtegskaqdlilgavllleqymaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSIGEMKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSGQVRL 143

Db 121 llgalqllgtqlpqrtrttahkdpnaiflqlgllrgkvrlflllvegptlcvrtppt 180
Qy 144 LLGALQLLGTQLPQGRTRTTAKDPNALFLSLIQQLLRGKVRFLLLVEGPTLCVRRTPPT 203

```

```

Db 181 avpsstsqlltlnkfmrtsqletnfsartagpqlslrlqgrfvkitpqdlnqtsrs 239
Qy 204 AVPSSTSQLTLNKFMRTSGLTWNFSVARTAGPGLSLRLQGRFVKITPQQLNQTSRS 263

Db 240 ldqipyvlnrhellngtrglfpgprrtllqadplsagtsadtslppnlqpygspephtp 299
Qy 264 PVQISGTLNTHGVPVNGTHGLFAGTSLQTLLEASDISPGAFNKGSLAFNLQGGLLPPSPILA 323

Db 300 pteqytlfpptlptp--vv-qlhpllpdpseaptptptapllntsythcmleaqe 352
Qy 324 PDGH-TTEPPSPALPTTHGSPQQLHPLFDPDTMPTNSTAPHVMTYMPHPRNLSEOT 378

RESULT 3
ENTRY A55530 #type complete
TITLE megakaryocyte growth and development factor, long form - human
ALTERNATE_NAMES MPL ligand, long form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Feb-1995; #sequence_revision 20-Feb-1995; #text_change 31-Mar-1995

ACCESSIONS A55530
REFERENCE #authors Chang, M.; McNinch, J.; Basu, R.; Shutter, J.; Hau, R.; Perkins, C.; Mar, V.; Suggs, S.; Welcher, A.; Li, L.; Lu, H.; Bartley, T.; Hunt, P.; Martin, F.; Samal, B.; Bogenberger, J.
#journal J. Biol. Chem. (1995) 270:511-514
#title Cloning and characterization of the human megakaryocyte growth and development factor (MGDF) gene.
#accession A55530
#status preliminary
#molecule_type DNA
#residues 1-286 #label CHA
#cross-references GB:U17071
#note sequence not compared to nucleotide translation

GENETICS
#gene MGDF
#map_position 3q26.3
KEYWORDS alternative splicing; cytokine
SUMMARY #length 286 #molecular-weight 31544 #checksum 6126

DB 9; Score 983; Match 83.6%; QryMatch 37.2%; Pred. No. 1.27e-113;
Matches 133; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

Db 1 meltlllvaamlavartlaspvaccdrllnkllrdshlhrlerlscpcvdpplapv 60
Qy 24 MELTDLLLAAMLLAVARLTSSVPACDPRLLNKLGRDHLHSLRLSCQPDVPLSIPV 83

Db 61 llpavdfsfgewktqtegskaqdlilgavllleqymaargqlpcslslqqlsgqvrl 120
Qy 84 LLPAVDFSIGEMKWTQTEQSKAQDILGAVSLLEGVMAARGQLEPSCLSLLGQLSGQVRL 143

Db 121 llgalqllgtqlpqrtrttahkdpnaiflqlgllrgkvrlflllvegptlcvrtppt 159
Qy 144 LLGALQLLGTQLPQGRTRTTAKDPNALFLSLIQQLLRGKVRFLLLVEGPTLCVRRTPPT 182

RESULT 4
ENTRY S28148 #type complete
TITLE erythropoietin - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change

```



```

FEATURE
1-27      #domain signal sequence #status Predicted #label SIG\
28-193    #product erythropoietin #status experimental #label MAT\
34-188,56-60 #disulfide bonds #status experimental\
51,65,110 #binding site carbohydrate (Asn) (covalent) #status
          experimental\
153       #binding site carbohydrate (Ser) (covalent) #status
          experimental
SUMMARY   #length 193 #molecular-weight 21307 #checksum 6371

DB 2; Score 131; Match 22.7%; QryMatch 5.0%; Pred. No. 2.41e-02;
Matches 22; Conservative 33; Mismatches 42; Indels 0; Gaps 0;

Db 12 LLLLSLPLGLPVGPAPPLICDRVLYLLEAKEANITGCAESLNNENITVPTD 71
      |||:|:|:| | | | | | | | | | | | | | | | | | | | | | |
Qy 29 LLLAAMLLAVARUTLSSVPAPACDPLRLKLLRDSHLHLSRLSQCPDVPDPLSIPVLLPAV 88
      ::|| | | | | | | | | | | | | | | | | | | | | | | | | |

Db 72 kvnfyawkrmevgqqavevwqglallseavlrgqall 108
      ::|| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 89 DFSLGEKKTQTEQSKAQDILGAVSLLEGVMAARGQL 125
      ::|| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
ENTRY   S24407 #type complete
TITLE   formin isoform IV - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     19-Feb-1994; #sequence_revision 19-Feb-1994; #text_change
        19-Feb-1994
ACCESSIONS S24407
REFERENCE  S24407
#authors   Jackson-Grusby, L.; Kuo, A.; Leder, P.
#journal    Genes Dev. (1992) 6:29-37
#title      A variant limb deformity transcript expressed in the
            embryonic mouse limb defines a novel formin.
#cross-references MUID:92112033
#accession  S24407
#status     preliminary
#residues   1-1206 ##label JAC
#cross-references EMBL:X62379
SUMMARY   #length 1206 #molecular-weight 133463 #checksum 501

DB 12; Score 117; Match 40.0%; QryMatch 4.4%; Pred. No. 5.28e-01;
Matches 20; Conservative 12; Mismatches 16; Indels 2; Gaps 2;

Db 658 IPPPPPLPP-GLGPLPAPPAPPVCPVSPPPPP-PPPTPVPPSGDPPP 705
      :|:|:|:|:| | | | | | | | | | | | | | | | | | | | |
Qy 316 LPSPSLAPDGHTPFPSPALPTHTGSPQQLHPLFPDPTTMTNSTAPHP 365
      ::|||:|:| | | | | | | | | | | | | | | | | | | | |

RESULT 8
ENTRY   S11515 #type complete
TITLE   formin - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     22-Jan-1994; #sequence_revision 22-Jan-1994; #text_change
        22-Jan-1994
ACCESSIONS S11515
REFERENCE  S11515
#authors   Woychik, R.P.; Maas, R.L.; Zeller, R.; Vogt, T.F.; Leder, P.
#journal    Nature (1990) 346:850-853
#title      'Formins': proteins deduced from the alternative transcripts
            of the limb deformity gene.
#accession  S11515
#status     preliminary
#residues   1-1468 ##label WQY

```

```

#cross-references EMBL:X53599
SUMMARY   #length 1468 #molecular-weight 163808 #checksum 6099

DB 12; Score 117; Match 40.0%; QryMatch 4.4%; Pred. No. 5.28e-01;
Matches 20; Conservative 12; Mismatches 16; Indels 2; Gaps 2;

Db 884 IPPPPPLPP-GLGPLPAPPAPPVCPVSPPPPP-PPPTPVPPSGDPPP 931
      :|:|:|:|:| | | | | | | | | | | | | | | | | | | | |
Qy 316 LPSPSLAPDGHTPFPSPALPTHTGSPQQLHPLFPDPTTMTNSTAPHP 365
      ::|||:|:| | | | | | | | | | | | | | | | | | | | |

RESULT 9
ENTRY   JQ0173 #type complete
TITLE   erythropoietin precursor - crab-eating macaque
ORGANISM #formal_name Macaca fascicularis #common_name crab-eating
          macaque
DATE     07-Sep-1990; #sequence_revision 07-Sep-1990 #text_change
        03-Mar-1995
ACCESSIONS JQ0173
REFERENCE  JQ0173
#authors   Lin, F.K.; Lin, C.H.; Lai, P.H.; Browne, J.K.; Egrie, J.C.;
            Smalling, R.; Fox, G.M.; Chen, K.K.; Castro, M.; Suggs, S.
#journal    Gene (1986) 44:201-209
#title      Monkey erythropoietin gene: cloning, expression and
            comparison with the human erythropoietin gene.
#cross-references MUID:87055236
#accession  JQ0173
#molecule_type mRNA
#residues   1-192 ##label LIN
#experimental_source kidney
COMMENT   This protein is the principal hormone involved in the regulation of
          erythrocyte differentiation and the maintenance of a
          physiological level of circulating erythrocyte mass.
CLASSIFICATION #superfamily erythropoietin
KEYWORDS        erythropoiesis; glycoprotein; hormone
FEATURE         1-24
                25-192
                51,65,110
                predicted
SUMMARY   #length 192 #molecular-weight 21113 #checksum 5284

DB 4; Score 114; Match 22.0%; QryMatch 4.3%; Pred. No. 9.99e-01;
Matches 22; Conservative 32; Mismatches 46; Indels 0; Gaps 0;

Db 12 LLLSLVPLGLPVGPAPPLICDRVLYLLEAKEANVTMGSCESLNNENITVPTD 71
      |||:|:|:| | | | | | | | | | | | | | | | | | | | | |
Qy 29 LLLAAMLLAVARUTLSSVPAPACDPLRLKLLRDSHLHLSRLSQCPDVPDPLSIPVLLPAV 88
      ::|||:|:| | | | | | | | | | | | | | | | | | | | |

Db 72 kvnfyawkrmevgqqavevwqglallseavlrgqavlans 111
      ::|| | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 89 DFSLGEKKTQTEQSKAQDILGAVSLLEGVMAARGQLPS 128
      ::|| | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
ENTRY   A24901 #type complete
TITLE   erythropoietin precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     19-May-1989; #sequence_revision 19-May-1989 #text_change
        18-Jun-1993
ACCESSIONS A24901
REFERENCE  A24901
#authors   McDonald, J.D.; Lin, F.K.; Goldwasser, E.
#journal    Mol. Cell. Biol. (1986) 6:842-848

```



```
#formal_name Homo sapiens #common_name man
#DATE 23-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
23-Mar-1993
ACCESSIONS S01387
REFERENCE S01387
authors Sillekens, P.T.G.; Beijer, R.P.; Habets, W.J.; van Venrooij,
W.J.
#journal Nucleic Acids Res. (1988) 16:8307-8321
#title Human U1 snRNP-specific C protein: complete cDNA and protein
sequence and identification of a multigene family in
mammals.
#cross-references MUID:88335591
#accession S01387
##molecule_type mRNA
##residues 1-159 ##label STL
##cross-references EMBL:X12517
SUMMARY #length 159 #molecular-weight 17394 #checksum 3819
DB 10; Score 113; Match 31.4%; OrigMatch 4.3%; Pred. No. 1.23e+00;
Matches 22; Conservative 18; Mismatches 27; Indels 3; Gaps 3;
Ddb 51 dktaaagqgkippptfeappgaamippsllpqpprpgmmmpachmgqpgpmnmggppp 110
| : || | : : : : : : : : : : : : : : : : : : : : : : : : : :
Qoy 297 DISPCATNKGSIAFNLCGLGPPSPSLADPGHT-PFPFSPAL-PTTH-GSPQQLHLPDP 353
| : || | : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb 111 pgmmvpvapap 120
| : || : || : ||
Qoy 354 STTNENSTAP 363
```

[illegible]

Search completed: Wed Sep 27 11:40:52 1995
Job time : 40 secs.

CC -/- SUBCELLULAR LOCATION: SECRETED.
 DR EMBL; L34169; WMTROA.
 DR PIR; S45330; S45330.
 KW GLYCOPROTEIN; HORMONE; SIGNAL.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 356 THROMBOPOEITIN.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 235 235 POTENTIAL.
 FT CARBOHYD 249 249 POTENTIAL.
 FT CARBOHYD 256 256 POTENTIAL.
 FT CARBOHYD 336 336 POTENTIAL.
 FT CARBOHYD 351 351 POTENTIAL.
 SQ SEQUENCE 356 AA; 37835 MW; 701294 CN;
 DB 7; Score 2482; Match 100.0%; QryMatch 93.9%; Pred. No. 0.00e+00;
 Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 meltclllaamlavaritispvapacdprlknklrdshllhstrlscqcdvdpisipv 60
 Qy 24 MELTDLLAAMLLAVARLTSSVPACDPRLLKLLRDSHLLHSLRLSQCPDVPDPLSPV 83
 Db 61 llpavdflsgewktqtegskaqdligavelllegvmaargqlpsclslqlglsqvrl 120
 Qy 84 LLPAVDFSLGEMKQTQESKAQDILGAVSLLLEGVMAARGQLPESCLSSLLGQLSQVRL 143
 Db 121 llgalqllgtqlpqrtaahkdpnaiflqlqllrgkvrfllllvegptlcvrtppt 180
 Qy 144 LLGALQGLLGTQLPQGRTHAKDPNALFLSLQQLLRGKVRFLLLVEGPTLCVRRTPPT 203
 Db 181 avpsstqltlmkfnrtsglltnfsvtargpqlrlgkvrflmvgstlcvrtppt 240
 Qy 204 AVPSSTQTLTKMKNFRTSGLLTNFVSVTARGPQLRLGKVRFLMVGSTLCVRRTPPT 263
 Db 241 pvqisgylnrthgpnvngthglfaqtqltleadspqafnkslafnlqglpsspsla 300
 Qy 264 PVQISGYLNRTHGPNVNGTHGLFAQTSLQTLLEASDISPGAFNKGSLAFNLQGLPSPSLA 323
 Db 301 pdghtpfpspalptthgspqllhplfpdptmpnstaphvmtymphnrlsqet 356
 Qy 324 PDGHTPFPPSPALPTTHGSPQLLHPLFPDPTMPNSTAPHVMTYMPHNRILSQET 379

RESULT 2
 ID TPO HUMAN STANDARD; PRT; 353 AA.
 AC P40225;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE THROMBOPOEITIN PRECURSOR (MEGAKARYOCYTE COLONY STIMULATING FACTOR)
 DE (C-MPL LIGAND) (ML) (MEGAKARYOCYTE GROWTH AND DEVELOPMENT FACTOR)
 DE (MGDF).
 GN THPO.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN (1)
 RP SEQUENCE FROM N.A.
 RM 94261202
 RA DE SAUVAGE F.J., DASS P.E., SPENCER S.D., MALLOY B.E., GURNEY A.L.,
 RA SPENCER S.A., DARGONNE W.C., HENZEL W.J., WONG S.C., KUANG W.-J.,
 RA OLES K.J., HUITGREN B., SOLBERG L.A. JR., GOEDEL D.V., EATON D.L.;
 RL NATURE 369:533-538 (1994).
 RN (2)

RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RM 94291201
 RA BARTLEY T.D., BOGENBERGER J., HUNT P., LI Y.S., LU H.S.,
 RA MARTIN F., CHANG M.S., SAMAL B.B., NICHOL S., BOSSELMAN R.A.;
 RL CELL 77:1117-1124 (1994).
 CC -/- FUNCTION: ACTS AS A CIRCULATING REGULATOR OF PLATELET NUMBERS.
 CC MEGAKARYOCYTIC LINEAGE SPECIFIC GROWTH AND DIFFERENTIATION
 CC FACTOR.
 CC -/- SUBCELLULAR LOCATION: SECRETED.
 DR EMBL; L33410; HSMCLMPL.
 DR EMBL; U11025; HS11025.
 DR PIR; S45331; S45331.
 DR MIM; 600044; 11TH EDITION.
 KW GLYCOPROTEIN; HORMONE; SIGNAL.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 353 THROMBOPOEITIN.
 FT CARBOHYD 197 197 POTENTIAL.
 FT CARBOHYD 206 206 POTENTIAL.
 FT CARBOHYD 234 234 POTENTIAL.
 FT CARBOHYD 255 255 POTENTIAL.
 FT CARBOHYD 340 340 POTENTIAL.
 FT CARBOHYD 348 348 POTENTIAL.
 FT CONFLICT 113 113 Q -> E (IN REF. 2).
 SQ SEQUENCE 353 AA; 37822 MW; 702923 CN;

DB 7; Score 1748; Match 71.6%; QryMatch 66.1%; Pred. No. 1.52e-287;
 Matches 255; Conservative 47; Mismatches 49; Indels 5; Gaps 4;
 Db 1 meltclllvmltlrtlssvpapacdprlrvlsklrdshvhlsharlsqcpvhlptpv 60
 Qy 24 MELTDLLAAMLLAVARLTSSVPACDPRLLKLLRDSHLLHSLRLSQCPDVPDPLSPV 83
 Db 61 llpavdflsgewktqtegskaqdligavelllegvmaargqlpsclslqlglsqvrl 120
 Qy 84 LLPAVDFSLGEMKQTQESKAQDILGAVSLLLEGVMAARGQLPESCLSSLLGQLSQVRL 143
 Db 121 llgalqllgtqlpqrtaahkdpnaiflqlqllrgkvrfllllvegptlcvrtppt 180
 Qy 144 LLGALQGLLGTQLPQGRTHAKDPNALFLSLQQLLRGKVRFLLLVEGPTLCVRRTPPT 203
 Db 181 avpsstqltlmkfnrtsglltnfsvtargpqlrlgkvrflmvgstlcvrtppt 239
 Qy 204 AVPSSTQTLTKMKNFRTSGLLTNFVSVTARGPQLRLGKVRFLMVGSTLCVRRTPPT 263
 Db 240 lqdpqynlrthgpnvngthglfaqtqltleadspqafnkslafnlqglpsspsla 299
 Qy 264 PVQISGYLNRTHGPNVNGTHGLFAQTSLQTLLEASDISPGAFNKGSLAFNLQGLPSPSLA 323
 Db 300 pdghtpfpspalptthgspqllhplfpdptmpnstaphvmtymphnrlsqet 352
 Qy 324 PDGHTPFPPSPALPTTHGSPQLLHPLFPDPTMPNSTAPHVMTYMPHNRILSQE 378

RESULT 3
 ID EPO MOUSE STANDARD; PRT; 192 AA.
 AC P07321;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE ERYTHROPOIETIN PRECURSOR.
 GN EPO.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

[illegible][illegible]

[illegible]

DR	EMBL; Z24681; OAERTPOIA.
KW	PROSITE; PS00817; EPO.
DR	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT	SIGNAL 1 27 BY SIMILARITY.
FT	CHAIN 28 194 ERYTHROPOIETIN.
FT	DISULFID 34 189 BY SIMILARITY.
FT	DISULFID 56 60 BY SIMILARITY.
FT	CARBOHYD 51 51 POTENTIAL.
FT	CARBOHYD 65 65 POTENTIAL.
FT	CARBOHYD 110 110 POTENTIAL.
SQ	SEQUENCE 194 AA; 21335 MW; 175348 CN;
DB	3; Score 128; Match 21.6%; OryMatch 4.8%; Pred. No. 2.03e-03;
Matches	21; Conservative 35; Mismatches 41; Indels 0; Gaps 0;
Db	12 llleflfplgvlqaprlldcervllyleareanatmgacgsfsenitvptd 71 : :: : :: : : : : : :
Qy	29 LLAAMLLAVARTLLSSVPADCPRLNKLRLHLLSHLSRUSQCDDVDFSLPIVLPAV 88 :: :: : :: : :
Db	72 kmfyakmvevgqqalewgglallseafgqall 108
Qy	89 DFLGCEWKQTQSKAQDILGAVSLLEGVMAARGQL 125
RESULT	9
ID	FORM MOUSE STANDARD; PRT; 1468 AA.
AC	Q05860;
DT	01-JUN-1994 (REL. 29, CREATED)
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE	FORMIN (LIMB DEFORMITY PROTEIN).
GN	LD.
OS	MUS MUSCULUS (MOUSE).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC	EUTHERIA; RODENTIA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=KIDNEY, AND TESTIS;
RM	90363291
RA	WOYCHIK R.P., MAAS R.J., ZELLER R., VOGT T.F., LEDER P.;
RL	NATURE 346:850-853(1990).
CC	-!- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC	AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
CC	INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.
CC	-!- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC	LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC	GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
CC	-!- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC	IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
CC	COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
CC	KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC	THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
CC	-!- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN
CC	ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A
CC	VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
CC	DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
CC	EMBL; X53599; MMULD.
DR	PIR; S11515; S11515.
DR	HSP; P19999; ICLG.
KW	NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
FT	DOMAIN 298 301 POLY-SER.
FT	DOMAIN 861 864 POLY-SER.
FT	DOMAIN 869 970 PRO-RICH.

FT	DOMAIN	977	981	POLY-SER.
SQ	SEQUENCE	1468 AA;	163809 MW;	11493196 CN;
DB	3;	Score	117;	Match 40.0%; QryMatch 4.4%; Pred. No. 4.14e-02;
	Matches	20;	Conservative	12; Mismatches 16; Indels 2; Gaps 2;
Db	884	ippppplpp-glqplppappipppcvspppppppp-pppctpvppsdgppp	931	
Qy	316	lppspslapdghtfppspalptthgspqqlhplfpdpsttmnstaphp	365	
		:	:	:
		:	:	:
RESULT	10			
ID	FOR4 MOUSE	STANDARD;	PRT;	1206 AA.
AC	Q05859;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	FORMIN 4 (LIMB DEFORMITY PROTEIN).			
GN	LD.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; RODENTIA.			
RC	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=EMBRYO;			
RM	92112033			
RA	GRUBSY-JACKSON L., KUO A., LEDER P.;			
RL	GENES DEV. 6:29-37(1992).			
CC	-!- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB			
CC	AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR			
CC	BE INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.			
CC	-!- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT			
CC	HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.			
CC	-!- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL			
CC	ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING			
CC	LIMB BUD.			
CC	-!- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE			
CC	PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN			
CC	SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE			
CC	TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS			
CC	DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR			
CC	DETERMINED.			
DR	EMBL; X62379; MMFOR.			
DR	PIR; S24407; S24407.			
DR	HSSP; P19999; 1CLG.			
KW	NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.			
FT	DOMAIN	635	638	POLY-SER.
FT	DOMAIN	644	744	PRO-RICH.
FT	DOMAIN	751	755	POLY-SER.
SQ	SEQUENCE	1206 AA;	133464 MW;	7742033 CN;
DB	3;	Score	117;	Match 40.0%; QryMatch 4.4%; Pred. No. 4.14e-02;
	Matches	20;	Conservative	12; Mismatches 16; Indels 2; Gaps 2;
Db	658	ippppplpp-glqplppappipppcvspppppppp-pppctpvppsdgppp	705	
Qy	316	lppspslapdghtfppspalptthgspqqlhplfpdpsttmnstaphp	365	
		:	:	:
		:	:	:
RESULT	11			
ID	EPO MACFA	STANDARD;	PRT;	192 AA.
AC	P07865;			
DT	01-AUG-1988 (REL. 08, CREATED)			
DT	01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)			

DT	01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE	ERYTHROPOIETIN PRECURSOR.
GN	EPO.
OS	MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC	EUTHERIA; PRIMATES.
RN	(1)
RP	SEQUENCE FROM N.A.
RM	87055236
RA	LIN F.-K., LIN C.-H., LAI P.-H., BROWNE J.K., EGRIE J.C., SMALLING R.,
RA	FOX G.M., CHEN K.K., CASTRO M., SUGGS S.;
RL	GENE 441:201-209(1986).
CC	-!- FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE
CC	REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A
CC	PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.
CC	-!- TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS
CC	AND BY LIVER OF FETAL OR NEONATAL MAMMALS.
CC	-!- SUBCELLULAR LOCATION: SECRETED.
DR	EMBL; M18189; MFEP.
DR	EMBL; M18188; MFEP.
DR	PIR; JQ0173; JQ0173.
DR	PROSITE; PS00817; EPO.
KW	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.
FT	SIGNAL 1 27 BY SIMILARITY.
FT	CHAIN 28 192 ERYTHROPOIETIN.
FT	DISULFID 34 187 BY SIMILARITY.
FT	DISULFID 56 60 BY SIMILARITY.
FT	CARBOHYD 51 51 BY SIMILARITY.
FT	CARBOHYD 65 65 BY SIMILARITY.
FT	CARBOHYD 110 110 BY SIMILARITY.
FT	CARBOHYD 152 152 BY SIMILARITY.
SQ	SEQUENCE 192 AA; 21113 MW; 175216 CN;
DB	3; Score 114; Match 22.0%; QryMatch 4.3%; Pred. No. 9.11e-02;
Matches	22; Conservative 32; Mismatches 46; Indels 0; Gaps 0;
Db	12 lllslvslplglvpagprlrcdsvrllylleakeanvtmgcsesclnehtlvdpdt 71 : : :: : :: : : : : :
Qy	29 LLTAAMLAVARLTLSPPVACPRLINKLRSHLLHSRLSQCPDPLSPVLIPAV 88 :: : : :: : : : : : :
Db	72 kmfyawkrmvgqqavewqglalleavirgqvavlans 111 :: : : :: : : : : : :
Qy	89 DFSIGEMKTQTQEKSQAQDILGAVSLILEGVMAARQLPEPS 128 :: : : :: : : : : : :
RESULT	12
ID	RUIC HUMAN STANDARD; PRT; 159 AA.
AC	P09234;
DT	01-MAR-1989 (REL. 10, CREATED)
DT	01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DE	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE	U1 SMALL NUCLEAR RIBONUCLEOPROTEIN C.
GN	SNRPC.
OS	HOMO SAPIENTS (HUMAN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC	EUTHERIA; PRIMATES.
RN	(1)
RP	SEQUENCE FROM N.A.
RM	88335591
RA	SILLEKENS P.T.G., BEIJER R.P., HABETS W.J., VAN VENROOIJ W.J.;
RL	NUCLEIC ACIDS RES. 16:8307-8321(1988).
RN	(2)
RP	SEQUENCE OF 25-131 FROM N.A.
RM	88088826

RA	YAMAMOTO K., MIURA H., MOROI Y., YOSHINOYA S., GOTO M., NISHIOKA K.,
RA	MIYAMOTO T.;
RL	J. IMMUNOL. 140:311-317(1988).
CC	-I- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U1.
CC	-I- SUBCELLULAR LOCATION: NUCLEAR.
DR	EMBL; X12517; HSUIRNP.C.
DR	EMBL; M18465; HSSNRNPA.
DR	PIR; S01387; S01387.
KW	NUCLEAR PROTEIN; RNA-BINDING; RIBONUCLEOPROTEIN; ZINC-FINGER.
FT	2N_FING 6 30 POTENTIAL.
FT	CONFLICT 25 27 CSG -> RSR (IN REF. 2).
FT	CONFLICT 94 98 APHMG -> TPIW (IN REF. 2).
FT	CONFLICT 101 101 P -> S (IN REF. 2).
FT	CONFLICT 129 131 HMP -> IQQ (IN REF. 2).
SQ	SEQUENCE 159 AA; 17394 MW; 144050 CN;

DB 6; Score 113; Match 31.4%; QryMatch 4.3%; Pred. No. 1.19e-01;
Matches 22; Conservative 18; Mismatches 27; Indels 3; Gaps 3;

Db 51 dkttaa fqqgki ptpfs apppagami pppps lpgpprpgmmppap hmgcggppmmppmgppp 110
 _ :
 Qv 297 DISPGAENKGSIAENLOGGIPPSLAPDGH-PPPPSPAL-PTTH-CSPPLHPLFPDP 353

```

Db      111  pgmmpvgpap 120
        :  || :  ||
Ov      354  STTMPNSTAP 363

```

RESULT 13

ID	EXTN TOBAC	STANDARD;	PRT;	620 AA.
AC	P13983;			
DT	01-JAN-1990	(REL. 13. CREATED)		
DT	01-JAN-1990	(REL. 13, LAST SEQUENCE UPDATE)		
DT	01-MAR-1992	(REL. 21, LAST ANNOTATION UPDATE)		
DE	EXTENSIN PRECURSOR	(CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).		
GN	HRCPTV73.			
OS	NICOTIANA TABACUM	(COMMON TOBACCO).		
OC	EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;			
OC	SOLANALES; SOLANACEAE.			
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. XANTHI; TISSUE=LEAF;			
RA	90128263			
RM	KELLER B., LAMB C.J.;			
RL	GENES DEV. 3:1639-1646(1989).			
CC	-1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN			
CC	THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE			
CC	MAIN ROOT.			
CC	-1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE			
CC	SE-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN			
CC	GLYCOSYLATED.			
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.			
DR	EMBL; X13885; NTEXT.			
DR	PIR; S06733; S06733.			
DR	HSP; P19999; ICLG.			
KW	REPEAT; CELL WALL; GLYCOPROTEIN; SIGNAL; STRUCTURAL PROTEIN;			
KW	HYDROXYLATION.			
		1	?	
FT	CHAIN	?	620	EXTENSIN.
FT	REPEAT	70	73	H-A-P-P.
FT	REPEAT	148	151	H-A-P-P.
FT	DOMAIN	229	242	2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT	REPEAT	229	235	1.
FT	REPEAT	236	242	2.

FT DOMAIN	205	620	CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN	499	600	3 X APPROXIMATE TANDEM REPEATS.
FT SEQUENCE	620 AA; 65406 MW; 2614500 CN;		
DB 3; Score	112; Match 33.3%; QryMatch 4.2%;	Pred. No. 1.55e-01;	
Matches	21; Conservative 16; Mismatches 23; Indels 3; Gaps 3;		
Db	369	lpppsppppp-pspfpppptyeagppppppaypplpapptyppppptyppppptyaapp	427
	: :	: :	: :
	: :	: :	: :
Qy	316	LPSPSLAPDGHPTPF-PPSPALPTTHGSPQQLHFPDPST-TMENSTAPHEVVTMYPHRP	373
Db	428	plp 430	
	:		
Ov	374	NLS 376	

RESULT	14
ID	GDF1 HUMAN
STANDARD;	PRT; 372 AA.

DT	01-AUG-1992 (REL. 23, CREATED)	
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)	
DT	01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)	
DE	EMERYONIC GROWTH FACTOR GDF-1 PRECURSOR.	
GN	GDF1.	
OS	HOMO SAPIENS (HUMAN).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	
OC	EUTHERIA; PRIMATES.	
[1]		
RN	SEQUENCE FROM N.A.	
RM	91239545	
RA	LEE S.J.;	
RL	PROC. NATL. ACAD. SCI. U.S.A. 88:4250-4254(1991).	
CC	-!- FUNCTION: GDF-1 MAY MEDIATE CELL DIFFERENTIATION EVENTS DURING EMERYONIC DEVELOPMENT.	
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN.	
CC	-!- SIMILARITY: TO OTHER GROWTH FACTORS OF THE TGF-BETA FAMILY.	
DR	EMBL; M62302; HSC61.	
DR	PIR: C39364; C39364.	
DR	PROSITE; PS00250; TGF BETA.	
KW	SIGNAL; GROWTH FACTOR; CYTOKINE; GLYCOPROTEIN.	
FT	SIGNAL 1 ?	POTENTIAL.
FT	PROPEP ?	POTENTIAL.
FT	CHAIN 254 372	EMERYONIC GROWTH FACTOR GDF-1.
FT	DISULFID 267 337	BY SIMILARITY.
FT	DISULFID 296 369	BY SIMILARITY.
FT	DISULFID 300 371	BY SIMILARITY.
FT	DISULFID 336 336	INTERCHAIN (BY SIMILARITY).
FT	CARBOHYD 206 206	POTENTIAL.
SQ	SEQUENCE 372 AA; 39502 MW; 618754 CN;	

DB 3; Score	108; Match 42.9%; QryMatch 4.1%; Pred. No. 4.32e-01;
Matches	21; Conservative 7; Mismatches 19; Indels 2; Gaps 2;

```

Db 14 llllllpls1pt-tapvppgaaallqalgrdepggapr1rpppv 61
    |||l:|l: || :||l: : || l||| :|| ||
Qv 29 llllaallavaritlssvpacpplinkl-lrduhllhslrslsogp 76

```

RESULT 15
ID GDA3 WHEAT STANDARD; PRT; 282 AA.

DT	13-AUG-1987 (REL. 05, CREATED)
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT	01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)

DE ALPHA/BETA-GLIADIN PRECURSOR (PROLAMIN) (CLASS A-III).
 OS TRITICUM AESTIVUM (WHEAT).
 * OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 * OC CYPERALES; GRAMINEAE.
 * RN (1)
 * RP SEQUENCE FROM N.A.
 RA 85234522
 RA OKITA T.W., CHEESBROUGH V., REEVES C.D.;
 RL J. BIOL. CHEM. 260:8203-8213(1985).
 CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -!- THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5 HOMOLOGY CLASSES.
 CC SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO SINGLE BASE
 CC SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS WITHIN OR NEAR
 CC DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF THE GENE
 CC FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
 CC EMBL; M11076; TAGLIABH.
 DR PIR; E22364; E22364.
 KW SEED STORAGE PROTEIN; REPEAT; SIGNAL; MULTIGENE FAMILY.
 FT SIGNAL 1 20
 FT CHAIN 21 282 ALPHA/BETA-GLIADIN A-III.
 SQ SEQUENCE 282 AA; 32236 MW; 403610 CN;

DB 3; Score 107; Match 32.8%; QryMatch 4.0%; Pred. No. 5.57e-01;
 Matches 22; Conservative 13; Mismatches 27; Indels 5; Gaps 5;

Db 55 qeafppqpy-ph-qpfpsqpyqpqpfpql-p-ypqtqpfppqpyqpqpqpqp 110
 | :|| : | : ||| : | : ||| | : | : ||| | : |

Qy 313 QGGLPSPSLAPDGHPTFFPSALPTTHGSPQLHPLFDPSTTMENSTAPHPVTMYPHP 372

Db 111 qpqisqq 117

: :||:

Qy 373 RN-LSQE 378

Search completed: Wed Sep 27 11:39:53 1995
 Job time : 22 secs.

c 22 20 1.9 427 57 T95155 ye39d06.r1 Homo sapie 2.41e-01
c 23 20 1.9 488 52 T77076 yd72h11.s1 Homo sapie 2.41e-01
24 20 1.9 302 32 RIC00791A Rice cDNA, partial se 2.41e-01
25 20 1.9 488 52 T75477 yd63b01.s1 Homo sapie 2.41e-01
c 26 20 1.9 236 36 T07633 EST05523 Homo sapiens 2.41e-01
27 20 1.9 359 3 CELK02603F C.elegans cDNA clone 2.41e-01
c 28 20 1.9 332 4 T46575 9838 Arabidopsis thal 2.41e-01
c 29 20 1.9 307 39 T27044 NIBT203H05R Homo sapi 2.41e-01
30 20 1.9 475 48 T62825 yf70b12.s1 Homo sapie 2.41e-01
c 31 20 1.9 414 18 T14376 yf82a06.r1 Homo sapie 2.41e-01
c 32 20 1.9 268 46 T51043 yb71b11.r1 Homo sapie 2.41e-01
c 33 20 1.9 361 18 R13977 yf68f04.r1 Homo sapie 2.41e-01
c 34 20 1.9 331 5 HSC098051 H. sapiens partial cD 2.41e-01
35 20 1.9 326 57 T95459 ye46g06.r1 Homo sapie 2.41e-01
36 20 1.9 325 13 M62087 EST00145 Homo sapiens 2.41e-01
37 20 1.9 426 58 T97685 ye54a02.s1 Homo sapie 2.41e-01
38 20 1.9 443 58 T99311 ye63b11.r1 Homo sapie 2.41e-01
39 20 1.9 300 57 T95103 ye44b08.s1 Homo sapie 2.41e-01
40 20 1.9 480 56 T90434 ye16h03.r1 Homo sapie 2.41e-01
c 41 20 1.9 323 29 R52051 yj71e09.s1 Homo sapie 2.41e-01
42 20 1.9 436 51 T71049 yc50g08.s1 Homo sapie 2.41e-01
43 20 1.9 562 44 T44094 7357 Arabidopsis thal 2.41e-01
44 20 1.9 290 38 T19301 e04006s Homo sapiens 2.41e-01
45 20 1.9 343 41 T31099 EST27367 Homo sapiens 2.41e-01

ALIGNMENTS

RESULT 1
LOCUS HSC1KF112 282 bp RNA EST 28-JAN-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1kf11.
ACCESSION F03047
KEYWORDS partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catasthini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 282)
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.
Genethon, B.P. 60, 91002 Evry Cedex France and Genetique
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801
Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE 2 (bases 1 to 282)
AUTHORS Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 282)
AUTHORS Auffray,C., Behar,G., Bois,F., Boucher,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: Integrated molecular analysis of the human genome and its
expression
C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
Cloning method: total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lsmid BA
vector;
Sequencing method: single read, full automatic;
Primer: (-21)/M13_universal;

cDNA sequence complementary to mRNA (3'end)
Stretch removed: removed at sequence 5'end
Normalization method: Bento Soares, P.N.A.S in press;
Genexpress_library_id: C;
Genexpress_sequence_id: aic-1kf11.

NCBI gi: 546604
FEATURES
source Location/Qualifiers
1..282
/organism="Homo sapiens"
/clone_lib="normalized infant brain cDNA from B.Souares,
Psychiatry Dept. Columbia University USA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/isolate="muscular atrophy patient"

BASE COUNT 84 a 93 c 48 g 53 t 4 others
ORIGIN
DB 7; Score 24; Match 66.7%; QryMatch 2.3%; Pred. No. 1.14e-05;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Db 86 ctccggagaccaccccccagccctgcgcacccctccctccgattccacctc 145
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 825 CTCAGGAACATCAGACACAGGCTCCTGCCACCAAGCTCCAGCTGATTCCTTC 884
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 ctgagccctacc 157
| |||||
Qy 885 CCCAACCCATCC 896

RESULT 2
LOCUS HSC1LH062 224 bp RNA EST 28-JAN-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1lh06.
ACCESSION F03088
KEYWORDS partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa/Eumycota group;
Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catasthini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 224)
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) to the EMBL/GenBank/DBJ databases.
Genethon, B.P. 60, 91002 Evry Cedex France and Genetique
Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801
Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE 2 (bases 1 to 224)
AUTHORS Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 224)
AUTHORS Auffray,C., Behar,G., Bois,F., Boucher,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: Integrated molecular analysis of the human genome and its
expression
C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
Cloning method: total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lsmid BA


```

Db 99 ctccggacaccgcccccaccctcccccaccctccctcccgatttccctcc 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
- Qy 825 CTCAGACATACACAGCGCTCCCTCCACCAACCTCCAGCTGGATATTCCTCTC 884

Db 159 ctgagaccatcc 170
   | |||||
- Qy 885 CCACACCATCC 896

RESULT 5
LOCUS R15865 433 bp mRNA EST 13-APR-1995
DEFINITION ya47a02.s2 Homo sapiens cDNA clone 53013 3'.
ACCESSION R15865
KEYWORDS EST.
SOURCE human clone=53013 library=Soares infant brain INIB vector=Lafmid BA
host=DH10B (ampicillin resistant) primer=21ml3 Rsite1=Not I
Rs1e2=Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGATTTCGGCGCGCGAGGAATTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 433)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Other ESTs: ya47a02.rl.exp
GDB: G00-425-949
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 393
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 768280
FEATURES
    source
        Location/Qualifiers
            1..433
                /organism="Homo sapiens"
                /clone="53013"
                /note="human"

BASE COUNT 116 a 57 c 113 g 142 t 5 others
ORIGIN

DB 19; Score 22; Match 86.7%; QryMatch 2.1%; Pred. No. 2.04e-03;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 189 tgtgtgagaagagcctcgagggttg 218
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Cp 1023 TGTGTTAGACAGCGCTGGTAGGGTGGG 994

RESULT 6
LOCUS T92900 255 bp mRNA EST 22-MAR-1995
DEFINITION ya23b11.s1 Homo sapiens cDNA clone 118557 3' similar to gb:M16660
HEAT SHOCK PROTEIN HSP 90-BETA (HUMAN);.
ACCESSION T92900
KEYWORDS EST.
SOURCE human clone=118557 library=Stratagene lung (#937210)
vector=pBluescript SK- host=SOER cells (kanamycin resistant)
primer=21ml3 Rsite1=EcoRI Rsite2=XhoI Normal lung tissue from a 72
year old male. Cloned unidirectionally. Primer: Oligo dT. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; 5' adaptor sequence:
5'-GAATTCGGACGAG-3'; 3' adaptor sequence:
5'-CTCAGATTTTTTTTTTTTTTTTTT-3'.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 255)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 125
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 724813
FEATURES
    source
        Location/Qualifiers
            1..255
                /organism="Homo sapiens"
                /clone="118557"
                /note="human"

BASE COUNT 85 a 53 c 61 g 49 t 7 others
ORIGIN

DB 57; Score 21; Match 85.7%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 166 caccagcagggggagccaggtggacag 193
   ||| ||||| ||||| ||| ||||| |||
Cp 85 CACAAGCAGGAGGAGCGCGCTGGACAG 58

RESULT 7
LOCUS R06417 337 bp mRNA EST 03-APR-1995
DEFINITION yf09a05.s1 Homo sapiens cDNA clone 126320 3' similar to gb:M23410
PLAKOGLOBIN (HUMAN);.
ACCESSION R06417
KEYWORDS EST.
SOURCE human clone=126320 library=Soares fetal liver spleen INFLS
vector=pTT3D (Pharmacia) with a modified polylinker host=DH10B

```

ORGANISM Homo sapiens
Eucaryotae: Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 434)
Hillier, L., Clark, N., Dubouche, T., Ellstrand, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Travaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 279
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

This clone is available royalty-free through LLNL ; contact the
Source: IMAGE Consortium, LLNL
IMAGE Consortium (info@image.llnl.gov) for further information.

```
source
1..434
/organism="Homo sapiens"
/clone="57051"
```

BASE COUNT	124 a	86 c	102 g	117 t	5 others
ORIGIN					

DB 50; Score 21; Match 76.9%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

RESULT	9				EST
LOCUS		R13972	344 bp	mRNA	
DEFINITION		vf68e06.r1	Homo sapiens	cdna clone	27404 5'.

DEL INFLUENZA
 ACCESSION R13972
 EST.
 KEYWORDS
 SOURCE human clone=27404 library=Soares infant brain INTB vector=Lafmid BA
 J100600111 NOMO sapiens CDNA clone 27404 3 .

SOURCE human clone=27404 library=Soares infant brain INIB vector=Lafmid BA host=DH10B (ampicillin resistant) primer=M13RPl Rsite1=Not I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGACAGATTCGGCCGCCGAGGATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the λmfmd BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 344)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Zannis-Hadjopoulos, D.

Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

GDB: G00-399-751
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 247
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

NCBI gi: 767048 Location/Qualifiers
1..344 /organism="Homo sapiens"
/clone="27404"
/note="human"

BASE COUNT 87 a 78 c 83 g 93 t 3 others

ORIGIN

DB 18; Score 21; Match 92.0%; QueryMatch 2.0%; Pred. No. 2.34e-02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 140 ctctttccaaaccactcctctac 164
||| ||||| ||||||| |||||||
Qy 878 CTCCTCCCAACCATCCTCTAC 902

RESULT 10 R31127 475 bp mRNA EST 28-APR-1995
LOCUS y61b05.rl Homo sapiens cDNA clone 134193 5' similar to contains 1L1
DEFINITION repetitive element i.
REMARKS R31127
KEYWORDS EST.
SOURCE human clone=134193 library=Soares placenta Nb2HP vector-ptT73D
(Pharmacia) with a modified polylinker host=DHI0B (ampicillin
resistant) primer=M13RP1 Reitel=Not I Reite2-Eco RI Female placenta
obtained at birth (full term). 1st strand cDNA was primed with a
Not I - oligo(dT) primer [5'
AACTCGAAGATTCCGCCGCGAGATTTTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Boudado.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 475)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 361
Source: IMAGE Consortium, LNLN

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium ([info@image.llnl.gov](#)) for further information.

NCBI gi: 786970
Location/Qualifiers
1..475
/organism="Homo sapiens"
/clone="J34193"
/note="human"

BASE COUNT 127 a 113 c 122 g 108 t 5 others
ORIGIN

DB 23; Score 21; Match 68.0%; QryMatch 2.0%; Pred.No. 2.34e-02;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Dd 303 agaagggcaaggacctccacgaagngnccccaccagaagcagcca 352
|| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cc 365 AGGAGCAGCGACTGTCCAGAAGCTGCCCGCAGGGATGAGGCACCA 316

RESULT 11
LOCUS R07547 388 bp mRNA EST 05-APR-1995
DEFINITION ye97e06.r1 Homo sapiens cDNA clone 125698 5'.
ACCESSION R07547
KEYWORDS EST.
SOURCE human clone=125698 library=Soares fetal liver spleen INFLS
vector=pT73D (Pharmacia) with a modified polylinker host=DH10B
(ampicillin resistant) primer=M13RP1 ReiteI=Pac I RsiteI=Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo(dT) primer [5'
AACTGGGAATTAAATAAAGATTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 388)
Hillier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,S., Tan,F.,
Trevaaskis,E., Waterston,R., Williams,A., Woldmann,P. and
Wilson,R.
The WashU-Merk EST Project
Unpublished (1995)

TITLE The WashU-Merk EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merk EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 303
Source: IMAGE Consortium, LNLN

This clone is available royalty-free through LNL ; contact the

	/note="human"			
BASE COUNT	82 a	103 c	121 g	101 t
6 others				

ORIGIN					
	DB 54;	Score	21; Match 74.4%; QryWatch 2.0%;	Pred. No.	2.34e-02;
	Matches	32; Conservative	0; Mismatches 11;	Indels	0; Gaps 0;
<hr/>					
Db	1	gaggagacaagtgaaggacctctccttgagcagggcccttc	43		
Qv	229	GAGGAGCAACAGGCACAGCATCTCTGGGAGCAGTACCCTTC	271		

RESULT	13					
LOCUS	HSC0JH031	285 bp	RNA	EST	05-NOV-1994	
DEFINITION	H. sapiens partial cDNA sequence: clone c-0703.					

DEFINITION H. sapiens partial cDNA sequence; clone c-03h1203 bp
ACCESSION Z42448
KEYWORDS partial cDNA sequence; transcribed sequence
SOURCE human.

REFERENCE
AUTHORS
Genexpress.
1 (bases 1 to 285)
Catarrhini; Homiidae; Homo.
Catarrhini; Homiidae; Homo.
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Metazoa; Eumetazoa; Bilateria; Cnidaria; Deuterostomia; Chordata; Eukaryota; mitochondrial eukaryotes; Metazoa/Eumycota group;

AUTHORS	Genexpress.
TITLE	Direct Submission
JOURNAL	Submitted (24-OCT-1994) to the EMBL/GenBank/DBJ databases. Genethon, B.P. 60, 91002 Evry Cedex France and Genetique Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE	2 (bases 1 to 285)
AUTHORS	Genexpress.
TITLE	The Genexpress cDNA program
JOURNAL	Unpublished

REFERENCE	3 (bases 1 to 285)
AUTHORS	Auffray, C., Behar, G., Bois, F., Boucher, C., da Silva, C., Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Marriage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani-Kabatchnik, C. and Tessier, A.
TITLE	IMAGE: Integrated molecular analysis of the human genome and its expression
JOURNAL	C.R. Acad. Sci., III, Sci. Vie 318, 263-272 (1995)
COMMENT	Clone library from B.Soares, Psychiatry Dept. Columbia University USA;

USA;

Cloning method: total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lacmid BA vector;

Sequencing method: single read, full automatic;

Primer: M13_reverse

cDNA sequence colinear to mRNA

Stretch removed: nothing

Normalization method: Bento Soares, P.N.A.S in press;

Genexpress_library_idt: C;

Genexpress_sequence_idt: y1c-0jh03;

No significant homology found with :
genbank release 81 swisssprot release 28.

```

NCBI gi: 565865
Location/Qualifiers
1..285
    /organism="Homo sapiens"

```

BASE COUNT	76 a	68 c	65 g	74 t	2 others				
ORIGIN	/dev_stage="3 months old" /isolate="muscular atrophy patient" /tissue_type="total brain" /clone_lib="normalized infant brain cDNA" /sex="Female"								
DB 6; Score	21; Match 92.0%; QryMatch 2.0%; Pred. No. 2,34e-07;								
Matches	23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;								
Db	141	ctctttcccaaacccatctctctac	165						
Qy	878	CTCCTTCCCAACCCATCTCTCTAC	902						
RESULT	14								
LOCUS	RICR0043A	230 bp	mRNA	EST	26-MAY-1995				
DEFINITION	Rice cDNA, partial sequence (R0043_1A).								
ACCESSION	D23738								
KEYWORDS	EST(expressed sequence tag).								
SOURCE	Oryza sativa (strain Nipponbare,) Seedling Root cDNA to mRNA.								
ORGANISM	Oryza sativa								
	Eukaryotes; mitochondrial eukaryotes; Chlorophyta/Embryophyta group; Charophyta/Embryophyta group; Embryophyta; Magnoliophyta; Lillipsida; Commelinidae; Poales; Poaceae; Oryza.								
REFERENCE	1 (bases 1 to 230)								
AUTHORS	Minobe,Y. and Sasaki,T.								
TITLE	Rice cDNA from root								
JOURNAL	Unpublished (1993)								
COMMENT	Submitted (2-NOV-1993) to DDBJ by:								
	Yuzo Minobe								
	Dept. Rice Genome Research Program								
	National Institute of Agrobiological Resources								
	Kannondai 2-1-2								
	Tsukuba, Ibaraki								
	Japan								
	Phone: 0298-38-7441								
	Fax: 0298-38-7468								
	PROJECT = 'RGP'.								

```

NCBI gi: 427605
Location/Qualifiers
1..230
/organism="Oryza sativa"
/strain="Nipponbare"
/dev stage="Seedling"
/sequenced_mol="cDNA to mRNA"
/tissue_type="Root"

BASE COUNT      67 a      73 c      34 g      51 t      5 others
ORIGIN

DB 33; Score      21; Match 74.4%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches          29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 53 cccacctccnccaacgccaacgcgcgcgcgcctcttt 91
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 978 cccttctgcttccaaacgccacgccacccctctcttct 1016

```

RESULT	15				
LOCUS	T87100	551 bp	mRNA	EST	17-MAR-1995
DEFINITION	_y08ae09.r1 Homo sapiens cDNA clone 115336 5' similar to gb:X53633 (HUMAN) ;.				

T87100
EST.
vector=pf73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13RPI Reitel=Pac I Rsite=Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 551)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 354
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

```

NCBI gi: 71542
FEATURES             Location/Qualifiers
     source            1..551
                        /organism="Homo sapiens"
                        /clone="115336"
                        /note="human"
BASE COUNT           108 a 137 c 151 g 152 t      3 others
ORIGIN
DB 55; Score         21; Match 80.6%; QryMatch 2.0%; Pred. No. 2.34e-02;
Matches             25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Dbb 451 ggggtnctgggtnctgctgaagggtacag 481
||||| ||| ||||| ||||| ||||| |||||
Cp 815 GGGGCTCTAGGCTGCTCGGTGAGGCTCCAG 785

Search completed: Wed Sep 27 23:48:17 1995
Job time : 252 secs.

```

 WAPSEPA

 (TH)

Release 2.1D John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
 Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 28 00:17:09 1995; MasPar time 61.85 Seconds
 876.152 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-252-491-18
 Description: (1:1062) from US08252491.seq
 Perfect Score: 1062
 N.A. Sequence: 1 ATGGAGCTGACTGATTCCT.....ATCTGTCAGAGCGTAA 1062
 Comp: TACCTCGACTGACTTACCA.....TAGACAGAGCTCTCCCAT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 61539 seqs, 25515148 bases x 2

Database: n-genseq
 1 n-gen1
 2 n-gen2
 3 n-gen3
 4 n-gen4
 5 n-gen5
 6 n-gen6
 7 n-gen7
 8 n-gen8
 9 n-gen9
 10 n-gen10
 11 n-gen11

Statistics: Mean 9.244; Variance 5.815; scale 1.590

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	84	7.9	1047	2	Q10572 Human Natriuretic Pep	4.67e-35
2	61	5.7	1047	2	Q10572 Human Natriuretic Pep	4.60e-21
3	44	4.1	91	9	Q51746 Oligonucleotide probe	3.04e-11
4	42	4.0	204	1	N81164 Base substituted E.co	3.91e-10

c	5	40	3.8	91	9	Q51746	Oligonucleotide probe	4.88e-09
c	6	40	3.8	204	1	N81164	Base substituted E.co	4.88e-09
c	7	29	2.7	565	6	Q35072	HCV envelope region n	2.60e-03
c	8	28	2.6	501	3	N50030	Sequence encoding new	7.98e-03
c	9	27	2.5	10596	9	Q51731	Plasmid pCiEBON for	2.41e-02
c	10	26	2.4	501	3	N50023	Sequence encoding new	7.17e-02
c	11	25	2.4	3871	2	N71302	HSV-1 gB and surround	2.09e-01
c	12	25	2.4	501	3	N50031	Sequence encoding new	2.09e-01
c	13	25	2.4	1956	10	Q58731	Encodes secretory alk	2.09e-01
c	14	24	2.3	565	6	Q35072	HCV envelope region n	5.98e-01
c	15	24	2.3	501	3	N50029	Sequence encoding new	5.98e-01
c	16	24	2.3	501	3	N50027	Sequence encoding new	5.98e-01
c	17	24	2.3	501	3	N50032	Sequence encoding new	5.98e-01
c	18	24	2.3	501	3	N50026	Sequence encoding new	5.98e-01
c	19	24	2.3	501	3	N50024	Sequence encoding new	5.98e-01
c	20	23	2.2	498	3	N50034	Sequence encoding new	1.67e+00
c	21	23	2.2	17041	3	Q21065	Genomic DNA of human	1.67e+00
c	22	23	2.2	501	3	N50028	Sequence encoding new	1.67e+00
c	23	22	2.1	920	5	Q32366	MAGE-10 genomic DNA.	4.57e+00
c	24	22	2.1	501	3	N50033	Sequence encoding new	4.57e+00
c	25	22	2.1	501	3	N50025	Sequence encoding new	4.57e+00
c	26	21	2.0	943	5	Q32356	MAGE-31 gene.	1.22e+01
c	27	21	2.0	1914	7	Q41701	Rat serotonin clone S	1.22e+01
c	28	21	2.0	2108	7	Q41702	Rat serotonin clone S	1.22e+01
c	29	20	1.9	2472	7	Q43967	Pokeweed antiviral pr	3.16e+01
c	30	20	1.9	3437	3	Q14937	Abelson Related Gene,	3.16e+01
c	31	20	1.9	7876	10	Q58707	E. coli multiple anti	3.16e+01
c	32	20	1.9	501	3	N50023	Sequence encoding new	3.16e+01
c	33	20	1.9	4481	1	N90958	Sequence encoding ger	3.16e+01
c	34	20	1.9	15672	2	Q10613	Rianodin receptor gen	3.16e+01
c	35	20	1.9	4487	3	Q20514	Encodes germ cell alk	3.16e+01
c	36	20	1.9	1409	3	N50474	Sequence of brain spe	3.16e+01
c	37	20	1.9	4258	3	Q22439	DNA of hgEco-1, encod	3.16e+01
c	38	20	1.9	12752	8	Q47380	Sequence which corres	3.16e+01
c	39	20	1.9	1237	2	N70314	Sequence encoding hum	3.16e+01
c	40	20	1.9	541	3	N50199	Sequence of cDNA clon	3.16e+01
c	41	20	1.9	3230	4	Q25595	Human Gnt I cDNA clon	3.16e+01
c	42	20	1.9	3871	2	N71302	HSV-1 gB and surround	3.16e+01
c	43	20	1.9	1265	6	Q36503	Sequence of zucchini	3.16e+01
c	44	20	1.9	325	8	Q59215	Human brain Expressed	3.16e+01
c	45	19	1.8	2226	5	Q32360	MAGE-5 cDNA.	7.93e+01

ALIGNMENTS

RESULT 1
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DT 09-APR-1991 (first entry)
 DE Human Natriuretic Peptide Receptor B.
 KW NPNB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
 KW hyperaldosteronism; glaucoma; guanyl cyclase.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal sequence
 FT Protein 12
 FT /label= mature NPNB
 FT Domain 23..455
 FT /label= extracellular domain
 FT /note= "binds natriuretic peptides A, B and C"
 FT Domain 456..456
 FT /label= transmembrane domain
 FT Domain 479..1047

```

FT /label= cytoplasmic domain
FT /note= "GC and protein kinase activity"
FT Modified -site 24..26
FT /label= N-glycos_site
FT Modified -site 35..37
FT /label= N-glycos_site
FT Modified -site 161..163
FT /label= N-glycos_site
FT Modified -site 195..197
FT /label= N-glycos_site
FT Modified -site 244..246
FT /label= N-glycos_site
FT Modified -site 277..279
FT /label= N-glycos_site
FT Modified -site 349..351
FT /label= N-glycos_site
FT Modified -site 600..602
FT /label= N-glycos_site
PN W09100292-A.
PD 10-JAN-1991.
PF 22-JUN-1989; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WP1; 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

DB 2; Score 84; Match 9.3%; QryMatch 7.9%; Pred. No. 4.67e-35;
Matches 89; Conservative 285; Mismatches 571; Indels 17; Gaps 17;

Db 76 synnannsavdknkyhdndnngvcvynaasvarnashwrnnnnntagavaasgskndh 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 1020 GTTTAGAGAGGGCTGTAGGGGTGGCGCTTGACAGAGAGGGTCAGGAGCAGGGGTG 961
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 136 yrtvrtqnsankngvntbqhwnwtaraamvndardtdrhnhyntngvnnanngsn 195
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 960 GAGCTGCACACAGGGTGGCGAGGTGGTGGAGAGGGAACAGCCGTATCTGTCCAGT 901
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 nevnhvyrnngunnathnnrangrvnyvcgnnnnnnnnnnnnnnnnnnnngd-yvyny 254
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 900 AGGAGGATGGTGGGAAGGAGGAATATCCAGGCTGGAGGTTGGGTGGCAGGCGCTGT 841
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 ndvngnsnrgntrkatgrndntrnnanraanntvntvyrnnnnnnnnnnnnnnnn 314
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 840 GTCTGATGTTCTCGAGAAATCTCCGGGGCTCTAGGGTCTCTGGCTGAGGGTCCAGGAA 781
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 nrarn-dhvgvngnsnnnnnagcnydgnnnyanvnnntnnngtrndgnrvnkmgrny 373
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 780 GAGTCCACAGAGTCCATTCAAGAGTGGTGTATCTCTGTCAGGTATCCGGGGATTGGTC 721
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 hqvtgvtvnmvdkndndrntdnvnmwngdhndghnnaabysganknnwtgrnnnnwvkann 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Cp 720 CAG-GGACCTGGAGGTTTGGTTTCAG-CAGACGAGCAATCT-TGGCTCTGAATCCCTGCTG 664
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 sdhnnncandhdnscdktnnetnnavngtgnntnmngvssnnnnrknmmnknnasmwr 493
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 663 CCACCTTCAGAGCCACAGGCGAGTAGTCTTGCTGAGGCGAGTGAAGTTTGTCTCCACAA 604
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 nrwnnnngnsnrhykgsdrtntnsrsgvsnmtahgkyknnnnantghnkgvnnvankhv 553
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 603 TCACAGAGTCTCTGTTGGGAGCTGCTTCACTGTGAGGACTAGACAGGTTCTCTGGGGAC 544
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 nk-krrnntnrwnnnnkhmrdrvnnhnrungacndnnnnnncvntvncnrgsnndnnnds 612
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 543 AGCTGTGGTGGGGGGGGGGGGCTGCAC-GCAGAGGGTGACCCCT-CCTA-CAAGCATCAG 487
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 nmndwmnrysnndvkvmmannhnsnshgsknsncvsvdrnvnkntdygnaenreta 672
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 486 GAAACGCCACCTT-TCCTCGGAGCAGGTGTGGAAGCTCAGGAGGATGAGAGCAAGTG- 312
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 ggtennndnnnnmnyannnnknvnnrttnaynnnnknkanannynnnnnhvvannnkrgntv 852
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 311 GGTCCAGAGTGTCCCGGCTGCGCATCAC-TCCCTCCACACAGAGGGTCACTGTCCAG 253
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 853 nanandsvntvnsdngvntansanstnmvntnnndhntcndannndndvkvntngday 912
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 252 AATGCTCTGCTGCTGCTGCTCCTCCATCTGGGTTTCCATTCTCCCAAGCTAAAGTCCAC 193
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 913 mvvegnngnrgnrhannmarannandavsnrnrhrhndnnrnrn-gvhtgnv-eagvv 970
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 192 AGCAGGACGACGACAGGTGTAGGCAAGGGTGAACCTCTGGGCATGGCTCAGTCTGCT 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 971 gnkmyrcnngdvtntasrmnsngnanknhvssttkdandnngcnmmnnrvgdmkqkqk 1030
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 132 CTGAGGAGCATGGGAGTCAACGACAGTTTACTGAGGACTCGGAGGTCAACAGCAGGAGG 73
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1031 mr 1032
: :
Cp 72 AG 71

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanylyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A, B and C]"

```



```

CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

DB 1; Score 40; Match 10.2%; QryMatch 3.8%; Pred. No. 4.88e-09;
Matches 10; Conservative 52; Mismatches 36; Indels 0; Gaps 0;

Db 93 hyrrmrnbvrdynrdaawcyccrvskvdcynachddhdyvbbvynvnhnnhn 152
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Cp 289 CCATCACTCCCTCCAGACAGAGGCTCACTCTCCAGAAATGCTGTGCTTGGCTCCT 230
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 cecbnnhvcnvhbnnhrnwayrhdzrdvhhccvch 190
   || : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 229 CCATCTGGGTTTTCATCTCCCAAGTAAAGTCGACA 192

RESULT 7
ID Q35072 standard; DNA; 565 BP.
AC Q35072;
DE HCV envelope region nucleic acid.
KW Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
KW non-A, non-B; amplify; ss.
OS Hepatitis C virus.
PN J04349885-A.
PD 04-DEC-1992.
PR 29-MAY-1991; JP-152169.
PA (TEIJ) TEIJIN LTD.
DR WPI; 93-022708/03.
PT Envelope region nucleic acid fragment - for type C hepatitis
   virus (1), for producing vaccine
PS Claim 1; Page 2; 13pp; Japanese.
CC This sequence encodes a novel envelope region of type C hepatitis
   virus (HCV). This fragment can be used for the preparation of a
   CC vaccine for hepatitis C. This fragment was prepared from the serum
   CC of non-A, non-B hepatitis patients and the envelope region DNA was
   CC amplified by PCR using the primer sequences given in Q35073-76.
SQ Sequence 565 BP; 61 A; 92 C; 106 G; 85 T;

DB 6; Score 29; Match 25.3%; QryMatch 2.7%; Pred. No. 2.60e-03;
Matches 22; Conservative 36; Mismatches 29; Indels 0; Gaps 0;

Db 296 cyrcyarraybyavvytyccrcbrygmrmwtwcbqbcayrtcgaytygtcythg 355
   |::| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 271 CTGCTGGAGGAGTGATGGCAGCAGGGGACAACTGGGACCCACTTGCCTCTCATCCCTC 330
   ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|

Db 356 ggrvgyrcyrcytcgtcscgytst 382
   ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
Qy 331 CTGGGGCAGCTTCTCGCAGAGTCGCT 357

RESULT 8
ID N50030 standard; DNA; 501 BP.
AC N50030;
DE Sequence encoding new modified human beta interferon polypeptides
   DE IFN-447.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..501
FT EP-163993-A.

```

```

PD 11-DEC-1985.
PR 17-MAY-1985; 105750.
PA (SEAR) SEARLE G D & CO.
PI Bell ID, Boseley PC, Porter AG;
DR WPI; 85-311944/50.
DR P-PSDB; P50029.
PT New modified human beta interferon polypeptide(s) - prepd. by
   PT plasmid transformed bacteria, with improved antiviral,
   PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2h, page 39; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
   CC INFs of the invention are more active and have different affinities
   CC for cell surface receptors (allowing selective targeting); they
   CC have higher therapeutic index; improved stability against microbial
   CC breakdown during synthesis; and better in vivo solubility and
   CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 110 A; 31 C; 69 G; 79 T;

DB 3; Score 28; Match 35.5%; QryMatch 2.6%; Pred. No. 7.98e-03;
Matches 33; Conservative 29; Mismatches 30; Indels 1; Gaps 1;

Db 45 ycartgcaraarytbybtgcarytbaaygngmdyrbqartaytygtbbaargaym- 103
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Cp 845 CCTGCTGTGATGTTCTCTCAGGAAATGTCGGGGCTCCTTAGGGTCTCGTGGGTCCA 786
   | : | | | : : : : : | : | : | : | : | : | : | : | : | : |

Db 104 gdcaygaytyymgdathcncargarttyg 136
   | : | | | : : : : : | : | : | : | : | : | : | : | : | : |
Cp 785 GGAAAGAGTCCACGAGTCCATTCAAGAGTTCCG 753

RESULT 9
ID Q51731 standard; DNA; 10596 BP.
AC Q51731;
DE 31-MAY-1994 (first entry)
DE Plasmid pCisEBON for subcloning huHGF variants.
KW Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT enhancer 1..611
FT /*tag= a
FT /note= "CMV enhancer/promoter"
FT promoter 758..775
FT /*tag= b
FT /label= SP6 promoter
FT /misc feature 845..849
FT /*tag= c
FT /note= "SP6 RNA start"
FT /misc feature 902..966
FT /*tag= d
FT /function= cloning_linker
FT /polyA signal 967..1107
FT /*tag= e
FT /note= "SV40 poly A"
FT /misc feature 1108..1531
FT /*tag= f
FT /function= SV40_origin
FT /misc feature 1580..4189
FT /*tag= g
FT /label= EBNA-1
FT /misc feature 4190..6374
FT /*tag= h

```



```

FT CDS 790..3498
FT /tag= h
FT /label= HSV-1 gB
FT /note= "Includes N-terminal hydrophobic leader and
FT a membrane-spanning sequence, a C-terminal
FT ionic sequence, and 9 N-linked
FT saccharide-addition sites"
FT 3'UTR 3499..3549
FT /tag= i
FT polyA_signal 3518..3525
FT /tag= j
FT polyA_site 3549..3549
FT /tag= k
FT misc_RNA 3549..3997
FT /tag= l
FT /note= "3' nonessential sequences to the BamHI
FT site"
FT site*
PN U54642333-A.
PD 10-FEB-1987.
PR 20-JUN-1984; 622496.
PR 16-SEP-1983; US-532996.
PR 20-JUN-1985; US-622496.
PA (PERS/) PERSON S.
PI Person S;
DR WPI; 87-056354/08.
PT Amino acid chain of glycoprotein B of HSV-1 and 2 - prepd. as
PT recombinant and used for vaccines for herpes simplex virus types 1
PT and 2.
PS Example; Table 1; 16pp; English.
CC !NOTE! This sequence has been indexed as represented in the
CC specification, except that bases 'E' have been replaced by 'N'.
CC The features have been indexed according to the legend of table 1 on
CC column 19/20 and the Sequence Summary of column 5 (sic). Note that
CC the base numbering of the features does not correspond to the
CC the sequence numbering below.
CC For another DNA sequence of HSB-1 gB see N711303 (P71135),
CC and for HSV-2 gB see N71399 (P71136).
CC A pure non-glycosylated amino acid (AA) chain comprising a sequence
CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
CC virus which is antigenic to HSV-1 of HSV-2, which contains no more
CC than 750 AA residues, and which includes AA residues 135-649
CC inclusive is claimed. It can be used to produce vaccines for
CC prophylaxis and treatment of HSV-1 and HSV-2.
SQ Sequence 3871 BP; 743 A; 1402 C; 544 G; 754 T;

DB 2; Score 25; Match 37.0%; QryMatch 2.4%; Pred. No. 2.09e-01;
Matches 51; Conservative 33; Mismatches 53; Indels 1; Gaps 1;

Db 2936 cbtbbggatcbacbtccggtbttbctcttcattcattccacacccctttctttbb 2995
Qy 846 CTCCTGCCACCCAACTCCAGCTGGATATTCCTCCCAACCCATCCTCTACTGG 905

Db 2996 -bgcbtgcbbctbttbctbttbctbctbctbctbctbctbctbctbctbct 3054
Qy 906 ACAGTATAGCTCTTCCTCTCCACCCAGCTTCGCCACCCCTGTGCTCCAGCTCCACC 965

Db 3055 acbtcatcbctbctbctbctbctbctbctbctbctbctbctbctbctbct 3072
Qy 966 CCTGCTCTCTGACCCCTTC 983

RESULT 12
ID N50031 standard; DNA; 501 BP.
AC N50031;

```

```

DT 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN 448.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..501
FT /tag= a
PN EP-163993-A.
PD 11-DEC-1985.
PR 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARIE G D & CO.
PI Bell LD, Boseley PC, Porter AG;
DR WPI; 85-311944/50.
DR P-PSDB; P50030.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2i, page 40; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targetting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 110 A; 30 C; 69 G; 80 T;

DB 3; Score 25; Match 32.8%; QryMatch 2.4%; Pred. No. 2.09e-01;
Matches 21; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

Db 45 ycartycaraarybybtggcarytbaaygmgydybgbartaytygtbaargaymg 104
Cp 845 CCGTGTCTGTATGCTCTCTCAGGAAATGTCGGGGCTCTCTAGGGTCTCGTGGGTCCA 786

Db 105 daar 108
Cp 785 GGAA 782

RESULT 13
ID Q58731 standard; cDNA; 1956 BP.
AC Q58731;
DT 29-SEP-1994 (first entry)
DE Encodes secretory alkaline phosphatase reporter protein.
KW Multicistronic expression unit; bicistronic vector system;
KW recombinant protein production; reporter protein; SEAP;
KW human placental alkaline phosphatase; ss.
OS Homo sapiens (placenta).
FH Key Location/Qualifiers
FT CDS 43..1563
FT /tag= a
FT /product= human SEAP
FT mat_peptide 94..1560
FT /tag= b
PN W09405785-A.
PD 17-MAR-1994.
PR 26-AUG-1993; E02294.
PR 27-AUG-1992; DE-228458.
PA (BEIE ) BEIERSDORF AG.
PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI Achterberg V, Dirks W, Dorschner A, Eichner W, Hauser H;
PI Meyer-Ingold W, Mielke H, Wirth M, Doerschner A;

```

DR WPI; 94-101190/12.
 PR P-PSDB; R50010.
 DT New multicistronic expression units - for producing equimolar
 PT amts. of polypeptide(s) in mammalian cells as hosts
 PS Example 1; Page 59-61; 109pp; German.
 CC Reporter genes coding for secretory alkaline phosphatase (SEAP)
 CC and for luciferase can be co-expressed using a bicistronic system.
 CC The SEAP is secreted due to the introduction of a stop codon at
 CC position 489 of the human placental AP sequence.
 SQ Sequence 1956 BP; 380 A; 659 C; 594 G; 333 T;

DB 10; Score 25; Match 79.1%; QryMatch 2.4%; Pred. No. 2.09e-01;
 Matches 34; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Dbl 1620 gccactgtccctgagtgcccgcctcctggggctcctgcttc 1662
 || ||||| ||||| ||||| ||||| ||||| ||||| ||
 Cp 267 GGTCACTGCTCCAGATGCTCTGGTCTTGCTCTCCTCATC 225

RESULT 14
 ID Q35072 standard; DNA; 565 BP.
 AC Q35072;
 DT 20-MAY-1993 (first entry)
 DE HCV envelope region nucleic acid.
 KW Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
 KW non-A, non-B; amplif; ss.
 OS Hepatitis C virus.
 PN J04349885-A.
 PD 04-DEC-1992.
 PF 29-MAY-1991; 152169.
 PR 29-MAY-1991; JP-152169.
 PA (TEIJ) TEIJIN LTD.
 DR WPI; 93-022708/03.

PT Envelope region nucleic acid fragment - for type C hepatitis
 PT virus (I), for producing vaccine
 PS Claim 1; Page 2; 13pp; Japanese.
 CC This sequence encodes a novel envelope region of type C hepatitis
 CC virus (HCV). This fragment can be used for the preparation of a
 CC vaccine for hepatitis C. This fragment was prepared from the serum
 CC of non-A, non-B hepatitis patients and the envelope region DNA was
 CC amplified by PCR using the primer sequences given in Q35073-76.
 SQ Sequence 565 BP; 61 A; 92 C; 106 G; 85 T;

DB 6; Score 24; Match 28.8%; QryMatch 2.3%; Pred. No. 5.98e-01;
 Matches 15; Conservative 23; Mismatches 14; Indels 0; Gaps 0;

Dbl 435 ctmbcmymgsmdyaygdgarrdyrcargytgyaaygtcdmtyacyc 486
 || ::||::||| |::||:: :||: :||: |:: :||:
 Cp 339 CTGCCCGAGGAGGCATGACAGGCAAGTTGGTCGCCATTGTCCCGCTGCTGCC 288

RESULT 15
 ID N50029 standard; DNA; 501 BP.
 AC N50029;
 DT 04-SEP-1991 (first entry)
 DE Sequence encoding new modified human beta interferon polypeptides
 DE INFN 446.
 KW Antiviral; cell growth regulator; immune system regulator;
 KW antiproliferative; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..501
 FT /*tag= a
 PN EP-163993-A.

[illegible]

 WAPSELEH (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
 Distribution rights by IntelliGenetics, Inc.

MPsrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 28 00:06:22 1995; MasPar time 567.91 Seconds
 1091.492 Million cell updates/sec

Tabular output not generated.

Title: >US-08-252-491-18
 Description: (1:1062) from US08252491.seq
 Perfect Score: 1062
 N.A. Sequence: 1 ATGGAGCTGACTGAATTGCT.....ATCTGTCTCAGGAGGGTAA 1062
 Comp: TACCTCGACTGACTTAACGA.....TAGACAGAGCTCTTCCCAAT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 267821 seqs, 291840210 bases x 2

Database: embl-new6

- 1 BCT
- 2 EST1
- 3 EST2
- 4 EST3
- 5 EST4
- 6 EST5
- 7 FUN
- 8 INV1
- 9 INV2
- 10 MAM
- 11 ORG
- 12 PIN
- 13 PRI
- 14 PRO
- 15 ROD
- 16 STS
- 17 SYN
- 18 UNC
- 19 VRT
- 20 VIR

Database:

- genbank89
- 21 BCT1
- 22 BCT2
- 23 BCT3
- 24 BCT4
- 25 BCT5

- 26 BCT6
- 27 INV1
- 28 INV2
- 29 INV3
- 30 INV4
- 31 INV5
- 32 MAM1
- 33 MAM2
- 34 PAT1
- 35 PAT2
- 36 PHG
- 37 PIN1
- 38 PIN2
- 39 PIN3
- 40 PIN4
- 41 PIN5
- 42 PIN6
- 43 PIN7
- 44 PRI1
- 45 PRI2
- 46 PRI3
- 47 PRI4
- 48 PRI5
- 49 PRI6
- 50 PRI7
- 51 PRI8
- 52 PRI9
- 53 ROD1
- 54 ROD2
- 55 ROD3
- 56 ROD4
- 57 ROD5
- 58 ROD6
- 59 ROD7
- 60 STR
- 61 STS1
- 62 STS2
- 63 STS3
- 64 STS4
- 65 SYN
- 66 UNA
- 67 VRL1
- 68 VRL2
- 69 VRL3
- 70 VRL4
- 71 VRL5
- 72 VRL6
- 73 VRT1
- 74 VRT2
- 75 VRT3

genbank-new6

- 76 BCT
- 77 EST1
- 78 EST2
- 79 EST3
- 80 EST4
- 81 EST5
- 82 EST6
- 83 EST7
- 84 EST8
- 85 INV
- 86 MAM
- 87 PHG


```

Qy 301 CAACCTGGGACCCACTTGCCTCTCATCCCTCCTGGGACAGCTTTCTTGACAGGTCCGCTCTC 360
Db 576 ctctctggggccctgcagagcctctcttggaaccagctctctccacaggcaggaccaca 635
|||||
Qy 361 CTCCTTGGGGCCCTGCAGAGCCTCTCTGGMACCCAGTCTCTCCACAGGGCAGGACACA 420
Db 636 gctcacaggatcccaatgccatctctctgagcttccaacactgctccagggaaaggtg 695
|||||
Qy 421 GCTCACAGGATCCCAATGCCATCTCTCTGAGCTTCCAACACACTGCTCCGAGCAAGGTG 480
Db 696 cgtttctgatgctgttaggaggtgccacctctgctcaggcggggcccccacccaccaca 755
|||||
Qy 481 CGTTTCTGATCTTCTAGAGGGTCCAACCTCTGCTCAGGGGGGCCCCACCCACACACA 540
Db 756 gctgtccccagcagaacctctctagtctccactgaacgagctcccaaacaggactct 815
|||||
Qy 541 GCTGTCCCGCAGCAGACCTCTCTAGTCTCTACACTGACAGCAGCTCCCAACAGGACTTCT 600
Db 816 ggattgtggagacaaacttcaactgctcagcagaactactggctctggcttctgaag 875
|||||
Qy 601 GCAATGTTGAGACAAACTTCACTGCTCAGCCAGACTACTGGCTCTGGGCTTCTGAAG 660
Db 876 tggcagcaggatccagagccagattcctgctgctgaacaaacctccagggtccctg 935
Qy 661 TGGCAGCAGGGATTTCAGGCCAAGATTCTGTGCTGTGCTGAACCAACCTCCAGGTCCCTG 720
Db 936 gaccaaatccccgataacctgaacagatatacgaactcttgaatggaactcgtgactc 995
|||||
Qy 721 GACCAAAATCCCGGATACCTGAACAGGATACAGAACTCTTGAATGGAACCTCGTGGACTC 780
Db 996 tttcctggacctcacgcaggacctaggagcccgagacatttctcaggaacatcacag 1055
|||||
Qy 781 TTTCTGTGACCTCAGCGAGGACCTAGAGGCCCGGACATTTCTCAGGACATCAGAC 840
Db 1055 acaggctccctgtccacccaacctccagctggatattctctctcccaacctcctcct 1115
|||||
Qy 841 ACAGGCTCCTGCGACCCACCACTCCAGCTCGATATTTCTTCTCCCAACCCATCCTCT 900
Db 1116 actgagcagatagctctctctctctccaccaccttgccacctctggtccagctc 1175
|||||
Qy 901 ACTGGACGATATAGCTCTCTCTCTCCACCCACCTTGCCCAACCCCTGTGGTCCAGCTC 960
Db 1176 caacctcgtctcctgacctctctgctcaacgcccaccctaccagcctctctctaaac 1235
|||||
Qy 961 CACCCCTCTCTCTGACCTCTCTGCTCCAAAGCCCAACCCCTACCGACCTCTCTTAAC 1020
Db 1236 acatctacaccactcccagatctgtctcaggaaggtaa 1277
|||||
Qy 1021 ACATCTTACACCCACTCCAGAAATCTCTCTCAGGAGGGTAA 1062
|||||

```

```

RESULT 3
LOCUS HSU11025 1341 bp mRNA PRI 15-OCT-1994
DEFINITION Human megakaryocyte growth and development factor (MGDF) mRNA,
complete cds.
ACCESSION U11025
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Eukaryote crown group; Metazoa/Eumycota
group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia;
Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii;
Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta;
Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 1341)
AUTHORS Bartley,T.D., Bogenberger,J., Hunt,P., Li,Y.S., Lu,H.S., Martin,F.,
Chang,M.S., Samal,B., Nichol,J.L., Swift,S. et.al.
TITLE Identification and cloning of a megakaryocyte growth and
development factor that is a ligand for the cytokine receptor Mpl
Cell 77 (7), 1117-1124 (1994)
JOURNAL
MEDLINE 94291201
REFERENCE 2 (bases 1 to 1341)
AUTHORS Samal,B.B.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1994) Babru B. Samal, Developmental Biology,
Angen Inc., Angen Center, Thousand Oaks, CA 91320, USA
COMMENT NCBI gi: 511223
FEATURES
source
1..1341
/clone="HFL-IVC7"
/organism="Homo sapiens"
/tissue_type="liver"
/dev_stage="fetal"
36..1097
/gene="MGDF"
/note="NCBI gi: 558078"
/codon_start=1
/product="megakaryocyte growth and development factor"
/translation="MELTELLIVMILLTARLTLSPPACDLRLSKILRDSVLIH
SRLSQCEVHPVLPVLLPAVDFSLGEMKQMEETKAQDILGAVTLLLEGVMAARQL
GPTCLSSLLGSGVRLILGALQLGTLQPPQRTTAHDPNAIFLSFOHLLRGV
REFIMVGSTLCVRRAPPTTAVPSRTSLVLTINELNRTSGLLETNFSARTTGSGL
LXKQGFRAKIPCLINOTBSLSDIPGYINRIHELINGTCLFLPGPSRRTLGAPDLS
GTSDTGSLPNLQPGYSPSPHTPGQYTLPLPTLPVQVQLHPLLPDSAPTPT
TSPLINTSYTHNSQLSQEG"
mat_peptide
99..1094
/gene="MGDF"
/product="megakaryocyte growth and development factor"
BASE COUNT 298 a 452 c 294 g 297 t
ORIGIN

```

```

DB 47; Score 1058; Match 99.8%; QryMatch 99.6%; Pred. No. 0.00e+00;
Matches 1060; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 36 atggagctgactgaattgctcctcgtggtcatgcttctcctaactcaaggctaacgctg 95
|||||
Qy 1 ATGGACTGACTGAAATGCTCTGCTGGTGTGCTATGCTTCTCTAACTGCAAGGCTAACGCTG 60
Db 96 tccagcccgctcctcctcgtgtgactccgagctcctcagtaaaactgcttctgactcc 155
|||||
Qy 61 TCCAGCCCGGCTCCTCTCTGCTGTGACCTCCGAGCTCCAGTAAACTGCTTCGTGACTCC 120
Db 156 catgtccttcaacagagactgagcagtgcccagaggttccacctttgctcaacctgtc 215
|||||
Qy 121 CATGTCTTCACAGCAGACTGACCCAGTGCCTCCAGAGGTTTCACCCCTTTGCTACACCTGC 180
Db 216 ctgctgctcgtgtggaacttagctgggagaaatggaaaccccatgatggagagacaag 275
|||||
Qy 181 CTGCTGCTGCTGTGGACTTTAGCTTGGGAGAAATGGAAACCCAGATGGAGGAGCAACAAG 240
Db 276 gcacaggacattctggagcagtgacctctctgctggaggagtgatggcagcacggga 335
|||||
Qy 241 GCACAGCAGATTCTGGGAGCAGTGAACCTTCTGCTGGAGGAGTGATGCCAGCAGGGGA 300
Db 336 caactgggaccacttgtctcatccctcctggggagcttctctgacagacctcctc 395
|||||
Qy 301 CAACTGGGACCCACTTGGCTCTATCCCTCTGGGGCAGCTTTCTTGGACAGGTCCTGCTC 360

```


Db 6784 caggaccctaggagcccgagcatttctcctcaggaacatcagacagagctccctgccacc 6843
 Qy 798 CAGGACCTAGAGCCCGGAGATTCTCTCAGAACATAGACAGGCTCCCTGCCACC 857
 Db 6844 caactccagcctggatttctctctcccaaccattctctactgagcagatatacgt 6903
 Qy 858 CAACCTCCAGCCTGGATTTCTCTTCCCAACCCATCTCTCTACTGAGCAGTACGCT 917
 Db 6904 ctctcctctccacacacattccacacccctgtgtccagctccacccctgtctctga 6963
 Qy 918 CTCTCTCTTCCACCCACCTTCCACCCCTGTGTGTCCAGCTCCACCCCTGTTCTGA 977
 Db 6964 cctctctgctcaacgcccacccctaccagcctctcttaaacacatcctaacaccactc 7023
 Qy 978 CCCTTCTGCTCCAGCCACCCCTTACAGCCCTCTCTTAACACATCTTACACCCATC 1037
 Db 7024 ccagaatctgtctcaggaaggtaa 7048
 Qy 1038 CCAGAATCTGTCTCAGGAGGCTAA 1062

RESULT 6
 LOCUS MUSTHROA 1486 bp mRNA ROD 06-JUL-1994
 DEFINITION Mus musculus thrombopoietin mRNA, complete cds.
 ACCESSION L34169
 KEYWORDS thrombopoietin.
 SOURCE Mus musculus cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
 Eutheria; Rodentia; Myomorpha; Muridae; Murinae.
 REFERENCE
 AUTHORS Lok S., Kaushansky, K., Holly, R.D., Kuijper, J.L., Lofton-Day C.E.,
 Oort, P.J., Grant, F.J., Heipel, M.D., Burkhead, S.K., Kramer, J.M.,
 Bell, L.A.N., Sprecher, C.A., Blumberg, H., Johnson, R., Prunkard, D.,
 Ching, A.F.T., Mathewes, S.L., Bailey, M.C., Forstrom, J.W.,
 Buddle, M.M., Osborne, S.G., Evans, S.J., Sheppard, P.O.,
 Presnell, S.R., O'Hara, P.J., Hagen, F.S., Roth, G.J. and Foster, D.C.
 TITLE Cloning and expression of murine thrombopoietin cDNA and
 stimulation of platelet production in vivo
 JOURNAL Nature 369, 565-568 (1994)
 MEDLINE 94261207
 COMMENT NCBI gi: 508540
 FEATURES
 source Location/Qualifiers
 1..1486
 /organism="Mus musculus"
 /sequenced_mol="cDNA to mRNA"
 174..1244
 /note="NCBI gi: 508541"
 /codon_start=1
 /function="c-mpl receptor ligand"
 /evidence=experimental
 /product="thrombopoietin"
 /translation="MELIDLLIAMLAVRLTLSSVPAPCDPRLINKLLRDSLHLH
 SRLSQCPDVPDLSIPVLPAPVDSIGEMKTOTQSKAQDILGAVSLILEGYMAARGOL
 EPGSLSSILQLSGQVRLLLGALQLLGTLQIPLOQRTTAHKDPNALFLSIQLLRGKV
 RFLIIVGEPITLCVRLPTTAVPSLTSLINKFPNRTSGLIETNF SVTARTAGFGL
 LSRLOQFVKITPQGLNQTSPVQISGYIARHGPVNGTHGIFAGTSLQTLASDIS
 PCAFNKGSIAPNLQGLPPSPSLAPDGHPTFPSPALPTTHGSPPLHPLPDPSTTM
 PNSTAPHPVTMHPHRLNLSQET"

BASE COUNT 331 a 484 c 337 g 334 t
 ORIGIN

DB 57; Score 602; Match 80.4%; QryMatch 56.7%; Pred. No. 0.00e+00;

Matches 856; Conservative 0; Mismatches 200; Indels 9; Gaps 4;

Db 174 atggagctgactgatttctcctggcgccatgcttcttgagtggaagactaactctcg 233
 Qy 1 ATGAGCTGACTGAATTGCTCTCTGCTGTCATGCTTCTCTTAAC TGAAGGCTAAACGCTG 60
 Db 234 tcaagcccgtagctcctgctgtgacccagagactcctaaataaaactgctgctgactcc 293
 Qy 61 TCCAGCCCGGGTCTCTCTGCTGTGACCTCCGAGCTCTCAGTAAACTGCTTCTGCTACTCC 120
 Db 294 caactccttcacagcagactgagtcagtcgacgcagtcgcagaccctttgtctatccctgtt 353
 Qy 121 CATGCTCTTCACAGCAGACTGACGACGTGCCAGAGGTTACCCCTTTGCTTACACCTGTC 180
 Db 354 ctgtcctctgctgtggaatttagcctgggagaatggaaacccacagcaggaagcaag 413
 Qy 181 CTGCTGCCCTGCTGTGACTTTAGCTTGGAGAGATGGAAA CCGAGATGGAGGAGACCAAG 240
 Db 414 gcacagacattctaggggcagtgctcctctactggaggagtgatggcagcagcagga 473
 Qy 241 GCACAGGACATTCTGGAGCAGTGACCCCTTCTGTGGAGGAGTGATGGCAGCACGGGA 300
 Db 474 cagtggaacccctcctgctctcatcctcctgggacagctttctgggcagggttcgcctc 533
 Qy 301 CAATGGGACCCACTTCCCTCTCATCCCTCTCTGGGCGAGCTTTCTGACAGAGCTCCGCTC 360
 Db 534 ctcttggggccctgcagggctccttaggaacccagcttctctacagggcaggaacaca 593
 Qy 361 CTCTTGGGGCCCTGCAGAGCTCTTGGAAACCCAGCTTCTCTCCACAGGGCAGGACCA 420
 Db 594 gctcaacagggcccccagctcctctcttctgagtgcaacaactgtctggggaaaggtcg 653
 Qy 421 GCTCACAGGATCCCAATCCATCTCTCTGAGCTTCCAACACCTGTCTCGAGGAAAGGTG 480
 Db 654 cgcttctcgtctgtgagaaggtccacacctctgtgcagcaggacctctgccaaacaca 713
 Qy 481 CTTTCTGTGATGCTGTGAGGAGGTCCACCCCTCTGGCTCAGGGGGGCCCCACCCACCA 540
 Db 714 gctgtcccaacagtagtacttccactcctcacactaaagaagttcccaacagagactct 773
 Qy 541 CCTGTCCCCACGACCACTCTTAGTCTCTCACTGACAGAGCTCCCAACACGAGCTTCT 600
 Db 774 ggattgtggagacgaacttcagtcacagccagaactgctggccttgagactctgagc 833
 Qy 601 GGAATTCTTGAGACAAACTTCACTGCTCAGCCAGAACTACTGCTCTGGGCTTCTGAG 660
 Db 834 aggttcaggagatttcagagattactcctcctggtcagctaaatcaaacctccaggtcc 893
 Qy 661 TGGCAGCAGGATTCAGAGCCCAAGATT---CCTGGTCTGCTGAACCAAACTCCAGGTCC 717
 Db 894 ccagtcctcaaatctctggataacctgaacggacacagcagctgtgaatggaactcatggg 953
 Qy 718 CTGACCAAAATCCCGGATATCCTGGAACAGGATACAGAACTCTTGAATGGAACCTCTGGA 777
 Db 954 ctcttctgctggaacctcacttcagacccctggaagcctcagacatctcgccggagatttc 1013
 Qy 778 CTCTTCTTGGACCCCTCAGCAGCAACCCCTAGGAGCCCGGACATTTCTCTCAGGAACATCA 837
 Db 1014 acaaaaggctccctggcattcaacctccaggggtgagcttctccttctccaagccttgc 1073
 Qy 838 GACAGAGGCTCCCTGCCCAACCTCCAGGCTGGATATTTCTCTTCCCAACCACTCCT 897
 Db 1074 cctgagtgacacacaccttccctcttcaactgcttgccttgccacacacacatggatctcca 1133

DB 57; Score 602; Match 80.4%; QryMatch 56.7%; Pred. No. 0.00e+00;

AUTHORS Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Sequin, C., Tuffnell, P.S. and Barrell, B.G.

TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome

JOURNAL Nature 310 (5974), 207-211 (1984)

MEDLINE 84270667

REFERENCE 2 (bases 1 to 172281)

AUTHORS Deininger, P.L., Bankier, A., Farrell, P., Baer, R. and Barrell, B.

TITLE Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome

JOURNAL J. Cell. Biochem. 19 (3), 267-274 (1982)

MEDLINE 83109311

REFERENCE 3 (bases 1 to 172281)

AUTHORS Farrell, P.J., Deininger, P.L., Bankier, A. and Barrell, B.

TITLE Homologous upstream sequences near Epstein-Barr virus promoters

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)

MEDLINE 83169725

REFERENCE 4 (bases 1 to 172281)

AUTHORS Farrell, P.J., Bankier, A.T., Sequin, C., Deininger, P.L. and Barrell, B.G.

TITLE Latent and lytic cycle promoters of Epstein-Barr virus

JOURNAL EMBO J. 2, 1331-1338 (1983)

REFERENCE 5 (bases 142687 to 159853)

AUTHORS Bankier, A.T., Deininger, P.L., Farrell, P.J. and Barrell, B.G.

TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus

JOURNAL Mol. Biol. Med. 1 (1), 21-45 (1983)

MEDLINE 85035713

REFERENCE 6 (bases 112620 to 125316)

AUTHORS Sequin, C., Farrell, P.J. and Barrell, B.G.

TITLE DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus

JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)

MEDLINE 85060424

REFERENCE 7 (bases 159853 to 172281)

AUTHORS Bankier, A.T., Deininger, P.L., Satchwell, S.C., Baer, R., Farrell, P.J. and Barrell, B.G.

TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences

JOURNAL Mol. Biol. Med. 1 (4), 425-445 (1983)

MEDLINE 85060428

REFERENCE 8 (bases 87650 to 92703)

AUTHORS Biggin, M., Farrell, P.J. and Barrell, B.G.

TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus

JOURNAL EMBO J. 3 (5), 1083-1090 (1984)

MEDLINE 84236104

REFERENCE 9 (bases 76089 to 79808)

AUTHORS Gibson, T., Stockwell, P., Ginsburg, M. and Barrell, B.

TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes

JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)

MEDLINE 84247360

REFERENCE 10 (bases 1 to 172281)

AUTHORS Hatfull, G.F., Barrell, B.G., Quinn, J. and McGeoch, D.

TITLE Unpublished

JOURNAL Unpublished

REFERENCE 11 (bases 1 to 172281)

AUTHORS Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and Griffin, B.E.

TITLE Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments

JOURNAL Nucleic Acids Res. 9 (13), 2999-3014 (1981)

MEDLINE 82014887

REFERENCE 12 (bases 1 to 172281)

AUTHORS Kozak, M.

TITLE Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes

JOURNAL Nucleic Acids Res. 9 (20), 5233-5262 (1981)

MEDLINE 82059504

REFERENCE 13 (bases 7315 to 9312)

AUTHORS Yates, J., Warren, N., Reisman, D. and Sugden, B.

TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)

MEDLINE 84222045

REFERENCE 14 (bases 45415 to 52824)

AUTHORS Jones, M.D., Foster, L., Sheedy, T. and Griffin, B.E.

TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus

JOURNAL EMBO J. 3 (4), 813-821 (1984)

MEDLINE 84207939

REFERENCE 15 (bases 45644 to 52450)

AUTHORS Jeang, K.T. and Hayward, S.D.

TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript

JOURNAL J. Virol. 48 (1), 135-148 (1983)

MEDLINE 83294686

REFERENCE 16 (bases 1 to 172281)

AUTHORS Farrell, P.J. and Barrell, B.G.

TITLE Direct Submission

JOURNAL Submitted (05-JUN-1984) to the EMBL/GenBank/DBJ databases

REFERENCE 17 (bases 1 to 172281)

AUTHORS Bodescot, M. and Perricaudet, M.

TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs

JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)

MEDLINE 87289053

REFERENCE 18 (bases 1 to 172281)

AUTHORS Laux, G., Perricaudet, M. and Farrell, P.J.

TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome

JOURNAL EMBO J. 7 (3), 769-774 (1988)

MEDLINE 88283646

REFERENCE 19 (bases 1 to 172281)

AUTHORS Farrell, P.J.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-1988) to the EMBL/GenBank/DBJ databases. Farrell, P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENT CDS

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAPF3 is the third leftward frame starting in Bam HI fragment A. BORF1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met

codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES of POLYA signals

This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAG is only listed when it is found in a position close to the end of a major reading frame.

SITES of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

Restriction enzyme sites.

Only the positions of the sites Bam HI (BAM) are listed.

RPT

This feature is used to define repetitive sequences.

SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORGPL

Denotes the region that encompasses an origin of replication (ori p). [13].

NUMBERING

The DNA sequence of B95-8 ERV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGCTGTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

NCBI gi: 59074

FEATURES	Location/Qualifiers
source	1..17281 /organism="Epstein-Barr virus" /strain="B95-8"
mRNA	58..272 /note="exon 2 terminal protein RNA"
mRNA	360..458 /note="exon 3 terminal protein RNA" complement(535) /note="polyA signal: AATAAA"
misc_feature	540..788 /note="exon 4 terminal protein RNA"
mRNA	871..951 /note="exon 5 terminal protein RNA"
mRNA	1026..1196 /note="exon 6 terminal protein RNA" complement(1192) /note="TATA: TATAAAT"
promoter	1280..1495 /note="exon 7 terminal protein RNA" complement(1383) /note="TATA: CATAAAA"
mRNA	1574..1682 /note="exon 8 terminal protein RNA"
promoter	1676 /note="TATA: TATAAAG"
promoter	1691 /note="TATA: TATTTAA BN-R1 late promoter before BNRFL, gives 4.1kb late RNA. Probably encodes non glycosylated 140kd protein in membrane antigen. Also two latent RNAs spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson et al, 1985). The longer one encodes terminal protein."
CDS	1736..5692

CDS

```

136...3972
/notes="BRRFL reading frame, 5 NKT/S; NCBI gi: 59075"
/codon start=1
/translation="MEERGRETQMPVARYGCPFLWRLFCQDCENIQEERLYELLSD
PSAIGLDGPGLTALNLVAGRTNNPQPHQERARELALVGLLNGQSGHGLGT
ESALEAGNNYVAGPDPWARSPTWSAEIQQLRLGLGLVPLVVMGQGFQFVHRS
RPSGFQFOALNHLYLFDAALPKYDSQVAAQGFQALLVAGETADTRPDLKLNWIF
GGRAAGQGLADELKTVALSDYLSGHVLIQVLTLDITMKVLSLRTVAHSLEHGIF
AAGTIOACMCPQLGEMRRHQGLFFFNVALTASSLGWYQTQTGADARAAARQQAQ
RAAEACHAKSGVPVAGPYRTINATLKGGEGQLTFWNGELGALIKHAQLDTPRDYGVH
YLYLMDGFPQMSGLTAPPCTYAESWSQAQVLTALFELSALYFAPCSIGYARQVPSA
VIEHLGSLVPRKGGLLTIALHLPDVKDQGLMGPMGASBPQGMQGFVSPFLNACSNV
FIFTRGEXKINGTFLQAALGRACMDACQOHVYLGSVPLGLLNFVNDLSASTAEM
MDDFSPVTFEFPFIOEGASSTVPILDYDEMSOISPTSLYPLHSLECSITLSHP
GSKEHLVLRHTDRVSGRRVQOQGVGLDPLADYSAFVHSSOVMTTRGGAPLPVRYMD
RMTXKLLVSAKPGGNNVSGTVTLTGQGYKYKSLOJRECTRILNMAEALLNAAQCPIL
DPEDLTLHLHLPDRPADNASWMEANTASQVADYARGVLYLTGASCPETGSSANP
MTVASVADKSPGGLITPVLOKLTGSLILNVRGCDQIGGSLFELQFLSDVATPTRA
PEALSXKLFRAVQOLKSGVILSHDLSOGELVTCLEVMALAGQGVRTIMPVASDY
LPMFAEHPGLVFEVEERSGEVLITKASNNYPAVLGRGVGQPMQFETQHQGVETV
LQSLRLILGTWSSFASEQEYCLRPDRNRSWMSVSDYGNALAVSPILTKNGSLPRLL
VTEPDRQVAVICAPGTRHESLLAATNAGCLLRFRFEVRNDTFDLKYVGLGIAIG
VGWGRDASAGRATVALNIFNPLADALIKLNRDPTFVVALGSEGVLAGLAVG

```

```

STNPPACGVENVQRSPILLIAPNASCMEFESRWLNISIPATTSVMIRGICRGCVLPCW
VQSGCIGQFTNLCMPYVYLQNAHQIACHFHSNGTDAMRFAMNYPNPTDQGNIAGLCS
RDCRHIALLCDPSLCTDFQWEHIPPAPCHPTGCSPTWLMFQAAHLWSLRHGRESE"
/notes="polyA signal: AATAAA"
3955
/notes="BAM: Bam H1 Nhet/h"
3994
/notes="BAM: Bam H1 h/C"
5408..5856
/notes="exon 9 terminal protein RNA"
5841
/notes="polyA signal: AATAAA, end of 4.1kb late RNA and TP
latent RNA."
5863
/notes="alternative end to TP cDNAs"
6097
/notes="TATA: TATAAGA"
6629..6795
/notes="Pol III RNA EBER 1"
complement(6823)
/notes="TATA: CATAAAT"
6956..7128
/notes="Pol III RNA EBER 2"
7315..9312
/notes="origin of replication, ori P (Yates et al., 1984,
1985)"
7421..8042
/notes="21x30bp repeats, binding sites for EBNA-1 (site I,
Rawlins et al., 1985). Tandem repeat part of orip (Reisman
et al., 1985). Also functions as a cell type specific
enhancer (Reisman et al., 1985; Lupton and Levine, 1985)"
7738
/notes="TATA: TATAAAT"
7888
/notes="TATA: TATAAAT"
8573
/notes="TATA: CATAAAT"
complement(8680)
/notes="polyA signal: AATAAA"
complement(8755)
/notes="polyA signal: AATAAA"
8962
/notes="polyA signal: AATAAA"
9021..9133
/notes="HPN: dyad symmetry, site II for EBNA-1 binding
(Rawlins et al., 1985). Dyad symmetry part of orip (Reisman
et al., 1985)"
complement(9398)
/notes="TATA: TATAAAT"
9631
/notes="TATA: TATAAAT BC-R1 late promoter before BCRF1"
9675..10187
/notes="BCRF1 reading frame; NCBI gi: 59076"
/codon start=1
/translation="MERRLVTVLQCLVLLYLAPECGGTQDCDNFQMLDLRDAFSRV
KTFQFKQVDNLLIKESILEDKFGYLGQALSEMIFQYLEEVPQAEQNQDPEAKDHY
NSIGENLKVTLRLRCHRFLLPCENKSKAVEQIKNAFNKLAQKGIYKAMSEDFINY
IEAYWTIKAR"
10076
/notes="TATA: GATAAAA"
complement(10148)
/notes="polyA signal: AATAAA"
10173

```

```

/notes="polyA signal: AATAAA"
10257
/notes="polyA signal: AATAAA, end of 0.8kb late RNA from
BCR1 and end of 1.6 kb late RNA, start unknown"
complement(10277)
/notes="polyA signal: AATAAA"
complement(10975)
/notes="TATA: CATAAAT"
11305
/notes="TATA: TACAAA; BCR2 promoter for highly spliced
EBNA latent RNAs."
11336..11480
/notes="exon C1 of Bodescot et al (1986) RNAs"
11524
/notes="TATA: TATAATT"
complement(11587)
/notes="polyA signal: AATAAA"
complement(11606)
/notes="TATA: CATAAAT"
11626..11657
/notes="exon C2 of Bodescot et al (1986) RNAs"
11796
/notes="TATA: TATAAGT"
complement(11799)
/notes="TATA: TATAAAA"
12001..15072
/notes="3072 repeat 1"
12541..13689
/notes="BCRF2 3072 repeat, reading frame 1"
13215
/notes="BAM: BamH1 C/W"
14352
/notes="TATA: TATAAAG BWR1 one of the promoters for highly
spliced EBNA and LP RNAs (Sample et al., 1986; Speck et al.,
1986)"
14384..14410
/notes="exon W0 of EBNA/LP RNAs"
14554..14619
/notes="exon W1 (also W66) part of leader protein (LP)
gene. LP is also called EBNA-5 (Dillner et al., 1986) and
EBNA-4 (Rowe et al., 1987)."
14559..14619
/notes="exon W1' (also W61) of EBNA/LP RNAs forms
initiator met when fused to exon W0 or exon C2."
14701..14832
/notes="exon W2 (also W132) part of LP gene"
15073..18144
/notes="3072 repeat 2"
15613..16761
/notes="BWRF1 reading frame 2"
16287
/notes="BAM: BamH1 W/W"
17424
/notes="TATA: TATAAAG"
17626..17691
/notes="Exon W1"
17773..17904
/notes="Exon W2"
18145..21216
/notes="3072 repeat 3"
18685..19833
/notes="BWRF1 reading frame 3"
19359
/notes="BAM: BamH1 W/W"

```

```

promoter      20496
/notes="TATA: TATAAAG"
mRNA          20698..20763
/notes="Exon W1"
mRNA          20845..20976
/notes="Exon W2"
repeat_region 21217..24288
/notes="3072 repeat 4"
CDS           21757..22905
/notes="BWRFL reading frame 4"
misc_feature  22431
/notes="BAM: BamH1 W/W"
promoter      23568
/notes="TATA: TATAAAG"
mRNA          23771..23835
/notes="Exon W1"
mRNA          23917..24048
/notes="Exon W2"
repeat_region 24289..27360
/notes="3072 repeat 5"
CDS           24829..25977
/notes="BWRFL reading frame 5"
misc_feature  25503
/notes="BAM: BamH1 W/W"
promoter      26640
/notes="TATA: TATAAAG"
mRNA          26842..26907
/notes="Exon W1"
mRNA          26989..27120
/notes="Exon W2"
repeat_region 27361..30432
/notes="3072 repeat 6"
CDS           27901..29049
/notes="BWRFL reading frame 6"
misc_feature  28575
/notes="BAM: BamH1 W/W"
promoter      29712
/notes="TATA: TATAAAG"
mRNA          29914..29979
/notes="Exon W1"
mRNA          30061..30192
/notes="Exon W2"
repeat_region 30433..33504
/notes="3072 repeat 7"
CDS           30973..32121
/notes="BWRFL reading frame 7"
misc_feature  31647
/notes="BAM: BamH1 W/W"
promoter      32784
/notes="TATA: TATAAAG"
mRNA          32986..33051
/notes="Exon W1"
mRNA          33133..33264
/notes="Exon W2"
repeat_region 33505..36576
/notes="3072 repeat 8"
CDS           34045..35193
/notes="BWRFL reading frame 8"
misc_feature  34719
/notes="BAM: BamH1 W/W"
promoter      35856
/notes="TATA: TATAAAG"
mRNA          36058..36123
/notes="Exon W1"

```

```

mRNA          36205..36336
/notes="Exon W2"
repeat_region 36577..39648
/notes="3072 repeat 9"
CDS           37117..38265
/notes="BWRFL reading frame 9"
misc_feature  37791
/notes="BAM: BamH1 W/W"
promoter      38928
/notes="TATA: TATAAAG"
mRNA          39130..39195
/notes="Exon W1"
mRNA          39277..39408
/notes="Exon W2"
repeat_region 39649..42720
/notes="3072 repeat 10"
CDS           40189..41337
/notes="BWRFL reading frame 10"
misc_feature  40863
/notes="BAM: BamH1 W/W"
promoter      42000
/notes="TATA: TATAAAG"
mRNA          42202..42267
/notes="Exon W1"
mRNA          42349..42480
/notes="Exon W2"
repeat_region 42721..45792
/notes="3072 repeat 11"
CDS           43261..44409
/notes="BWRFL reading frame 11"
misc_feature  43935
/notes="BAM: BamH1 W/W"
promoter      45072
/notes="TATA: TATAAAG"
mRNA          45274..45339
/notes="Exon W1"
misc_feature  45415..52824
/notes="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA          45421..45552
/notes="Exon W2"
misc_feature  45644..52450
/notes="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region 45793..47643
/notes="3072 repeat 12"
CDS           46333..47481
/notes="BWRFL reading frame 12"
misc_feature  47007
/notes="BAM: BamH1 W/Y"
mRNA          47761..47793
/notes="Exon Y1 Bodescot et al, 1984"
promoter      47831
/notes="TATA: TATAAGT"
mRNA          47878..47999
/notes="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
Strominger,1985), last common exon"
misc_feature  complement(48023)
/notes="polyA signal: AATAAA"
mRNA          48386..48444
/notes="exon Bodescot et al, 1984"
CDS           48386..50032
/notes="Coding exon for EBNA-2 (Sample et al,1986)"
CDS           48429..49964
/notes="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984;
Dillner et al, 1984)"

```

```

repeat_region 48678..48800
/note="14 x CCCCACCA repeats"
misc_feature 48848
/note="BAM: BamH1 Y/H"
promoter 49350
/note="TATA: TATAACA"
complement(49353)
/note="TATA: TATAAAA"
repeat_region 49525..49578
/note="9 x GGGCCA repeats"
49525..50032
/note="exon (Bodescot et al 1984)"
50033
/note="polyA signal: AATAAA, end of Bodescot T1 RNA and EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
complement(50156)
/note="TATA: TATAACT"
complement(50317)
/note="polyA signal: AATAAA, end of 2.5kb early RNA from 52817"
50578..52115
/note="12 x 125bp repeats"
complement(50578..52557)
/note="BHLF1 early reading frame"
52654..53697
/note="region homologous to Eco RI C of Raji"
complement(52817)
/note="TATA: GATAAAA promoter for 2.5kb early RNA containing BHLF1 (Jeang and Hayward, 1983; Freese et al, 1983)"
53759
/note="TATA: TATTAAAC likely promoter for class III and IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
53895
/note="DONOR: CCGGTAAC donor for splice to 54335 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
54335
/note="ACCEPT: TTTTCTAG acceptor from 48444 in class I, 47999 in class II, and 53895 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
54376..54948
/note="BHRF1 reading frame, limited homology to bcl-2 gene. Early gene in B95-8 cells and part of restricted EA complex."
54591
/note="TATA: TATAACA"
complement(54594)
/note="TATA: TATAAAT"
54853
/note="BAM: BamH1 H/F"
complement(54929)
/note="polyA signal: AATAAA"
complement(54977)
/note="TATA: TATAAAG"
55518
/note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb and 0.6kb early RNAs"
complement(55982..56935)
/note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27 in VZV and HFRF2 in CMV"
complement(55990)
/note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb early RNAs from 58568 and 57081"
complement(56132)

```

```

CDS
/note="TATA: TATAAAG"
complement(56948..58525)
/note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in VZV and HFRF1 in CMV; NCBI gi: 832975"
/codon_start=1
/translation="MAHKVTSANEPNPLTKGRLLSSCPVTRSGVTEVAOIGRTPKMED
FVPTVNDLKSQFEAVGLIMVLPANAEGIAYPLVHTYESLSPASTCRVCDLLD
FLVNHSDAPVAFFEDYALLCYCINAPRAWISLITGMDFLHLIKYFPWAGGLDSLF
MPSRILAIDIQLHFYICRCELPVSSSDMIRNANLGYKLEFLKSILTCGSPANFCFS
MPRTTFTLTLPGPRTCSQDVPCDVGRLYTALCCHLPITNRVQHPFLRAKGGGL
SPEITTRADYCGLLGTWQTDLLGGPGHAGIAEYSGDELAELAIAITRPEAGDH
SQGPCILAPMFLGRHKNASRTICLCSGLGAPHDADKTLDRFKSLILDSFGNMIKIL
RIVFLIKONTLLDVPCLRAWLQMCPTQDHFHFLCDPLCAINHSTNPSPVFCQI
YPPSFQAFRAALAAQNLQGVCDLSLTVLYFKSTQVARVCKTILNDVTKELDVLRL
IHGIDLVSQTSQVYV"
complement(57081)
promoter
/note="TATA: TATTAA before BFLF2; BFL2 promoter gives 1.1kb early RNA"
complement(58088)
promoter
/note="TATA: GATAAAA"
complement(58568)
promoter
/note="TATA: TATTAA before BFLF1, BFL1 promoter gives 2.3kb early RNA"
58832
/note="TATA: TATAAAA before BFRF1"
58891..59901
/note="BFRF1 early reading frame, 1 NXT/S, homologous to HFLF4 in CMV; NCBI gi: 832976"
/codon_start=1
/translation="MASPEERLLDELNNVIVFLCDSGLSEVERSCGAVHFRSGSQP
LCTVKLRHGQIVHLEFYKFLAFKLCRCNPSPVVISNNGIATTLRCFLHPEGLR
SQGSCPLGLSTDVDPKNSILMLGQDDFIKFKSPVPAEDLLKSMVCRAYITEH
RTMQFLVFOANAACKASRVMDMISDSQSLRSQVEDTGARVTCGGCPGPTGSHG
CLGSHVFRGGGWDLDNFSEATEDEASYPARDKDSMSESAAPKKELVRHPIRRH
RTRTRMRGSHSRVHVPETRETIVGGMARYSWRATPYLARVLAVTAVALLMFLR
W"
59610..61580
/note="BFRF2 early reading frame, homologous to HFLF5 in CMV"
complement(61062)
promoter
/note="TATA: GATAAAA"
61344
/note="TATA: TATTAA before BFRF3"
61456..62034
/note="BFRF3 early reading frame"
complement(62068)
misc_feature
/note="polyA signal: AATAAA"
62069
/misc_feature
/note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4, 3.1, 2.5 and 0.8kb early RNAs"
complement(62078..71527)
CDS
/note="BFLF1 reading frame, 1 NXT/S, analogous to VZV RF22; NCBI gi: 832977"
/codon_start=1
/translation="MSNGDWGQSORTGCPVRGIRTMVNPAGGSGGSAIRILCTA
SNQNHCKGFRFAGLCQVSNCLVLYKSFACRPLTSPPELDEIDECARLDMRQS
GLTKGHMMAQLDVPFSSVLRGGRVHIYRSAEIFGLVFLPAQIANSAVQSLAEVLR
GSYNGVAQFTLYICDIYAGAIITIEDGSFYLDPHCKOKDAAPCTPAHVRVSTVAHDIL
QVGAPCAQYTCVHLFLPEATEETDIPFMLEHYGVYDFEANGSGCDFLVGELVSS
DGEAAGTPGADSSPVMVLFERRIIPYNLRPLPSRSFTSDSPFAARYSPAKTNSPSS
PASAAPASAPASAPASAPASAPASAPASAPASAPASAPSPPLFIPICGLGHTP
GVPASTPPPASSGAAPQTPKCKGLCKDSPHKKPTSCRRILPLSSTTTEDDQLRTH
VPPHPPSAAARLPPEVIPPHQSPASPTPHPAVSTIAPSVTSPRLPILQIPILPQ
AAPSNNKIPILTPSPSTAAAPATTTTILSPPTTQQPPQSAAPAPSPLLPQQQPTPSA

```

```

APAPSLPQQPPSPSARAPSLPQQQP LP SATPAPPAQQLPSANTILEPKHP
PAADRAGTETSPSPFGQSPFGDASGGLVRLYSLDEEPLFSMSDEAEEDLAS
DIPITTEDEDMVEFNSLESGSAPTSPTITLDARSQYTOTTFDIETPMDFVPLE
SNIRIAGHTYQEQAIIVDPASNRNVEPADALSHDIYLLVTVLEQGLIRSDRSSVL
NILEEFKOWSGHLOVPTLDJEO LITSEINLIQNIANMSENKRGEGFHKHAKLEAC
LPSLATQAVRVDAGAKMAIEIQLAESDDGKFDLEARRLLDTLLSGDQOAGEGGG
EPEDNSIYRGPHVDPVPLVLDDESCKLLSLAEAAVAVARQOAGDEEDVRELLALTA
IETGAPPAVDPVPHINAVRSNAALHRECTADIRUKVASAASDIYSYLEDSPT
VMDFDLTHLRITCOIITASLP LINIRYTESIEMDYRELLYLGCTALSDMSGIPALERV
EEDPSIAPPEETVANKOKELETTRENEKRLTILDD IEAMGLIAGVASAPCAPIS
PASPSATPANIDPEATPLADTALITPIVEKYIANAGSIVGAKNPPTIIRUTOTIQ
QIVRSKYIMNLIKSTIYTDNY IASFEES IDHRLDPLDPEVQDIDRLIDPMV
SEALHTECMGNLITLEPARIVAQNFATSTIKETAANVLLPGLIAYVDATITGOAP
EDARLLSGQLQNSOTLLPKLKKFELSYLQKLNNDQIQKQEVQWRLNEGFK
PATTEQLEAFIDTAPNKEIKROYKRLQIMETGRKEKELREQEDKEKERRRAN
EAWERIKALGARPEPATSPDDWNTLLASLLPNDTSAAMAAAVARNTDILDSLTQ
ILAMLLGITVTRERELIASLLVDDGGAEREAEPGWFTDIETGFLARLAMPATPA
ATAEGGGGCGAEEAGALFARTAADAIIRSAQAOTROALQSDPKSAVNTDLEAPY
AEYERGLAGLEKRRAAEAALTAIVSEYVDTLPLEATNDPGQANLPPTIPOATAP
RILASDIALPKPKQLITRREDDLQATGDFSELLITEAAAEVRALEEQVRESQTIM
AKAHEMAASTRRGFTALVAVLSRSHDEAPDELRLSLIPSPKAPQAPLEAALHAA
LLEVEGMAQNSMDGMEELRLALATLDPKRVAGCKETVADWKRLSLRLAIIQEAQES
QIAGTQLQDVTQARGHTDPROJLIVVEARGALGASQYALIKDKLIRYASAKQS
AGNCSYRKSIAAKIRGI CEAVRGLSECALALAGAGAWINLAAADGEIHELTR
LLEVEGMAQNSMDGMEELRLALATLDPKRVAGCKETVADWKRLSLRLAIIQEAQES
QIAGTQLQDVTQARGHTDPROJLIVVEARGALGASQYALIKDKLIRYASAKQS
LWLETSPDCLMPA VVSADTOEPLNYIPVYHNFLEYVMPVLENPEAFSLTPAGPQ
AIGFQDDQERRRTLASVASARLSAAMADSYMDWTPVESNAGELLREIYVAPKALM
EDLADNPITVAMTLLAHASLIASRNHPYPAPATDREVILLQREEMALLVGTHPAYAA
AFLGAPSYAGLGSALARGDGLDLSLVLYRLVRSASGRGCMPESTTRCSNDG
EDARRLTRIAGCTPGIIFQDANEEMDTAALWHPHEFLGLVHNSSTARACMLL
LARGCPAFELQALHHSIRPLEGCVAFQDYLRDFVKQAYTRGEEILPRAEGLEVPRETP
SSYCTVTGRLRLNLPYGTPTGPKRSGDTPVSVFEAAVAAFLGRPLTLFVSQY
LENKLTQLQVRVAPLIXCDGSEPPERSIVETISINFLQDLDGYSSEPEPMSIFARQ
AWMLRELTENARAQKCARPTIVAILANRKNIIWKCFYIRHNLDPDQYFNAGASR
WPTDVLNPFYEHEDPPLFVGQYQLPNPNRYOELFSGFPVRVGHVSGDGFQSDNT
PASSDRIQLGGEGTDOEGKSTTAESEASGPPSPQSLLEKVPAGRPDMLSPTSSP
RDVTVPGLAAPTITLPGRLMARPYGAETRASESPPDRSGSPSPRPWKUSLELIPQ
APQDQSSPMWASQGP IVYTLSPHSTPASGSKKHTIQIQLVPSQKSPYPSAPY
PKQSGTGGIAPTSAASLTITGLQPDQTOASSODPPYGHSTIMOREKKQGGREEAEI
RPSATRLTAVGLRPNPVPVAGAAASATPAFDGENSGPP IQAPALGSLAAPAH
TPYGALAPRQKTAQRPQDAALPTPTIKAVGARVPVKATCALAAGARPQCTAAP
PSAASPPRVSLPVRSQOQSPAIP LPMHSGSEP GARPEVRLSQYRHAGCPOYTVRKE
APPSAASQLPMKPKCDSNYPYPSGSARYAPAFQALSFQSQVASPAPSSDQTLIMNT
PSYVTQFLSFTIIRVYVTGGTSGDLVVPVSGSPSSLSSTAPEQDLRLYSLTQSASRV
LSRFVSQLRRKRLERSHRLIADLERLKFYL"
62249
/notes="BAM: BamH1 F/Q"
62430..62477
/notes="Site III for EBNA-1 binding (Rawlins et al, 1985)"
66121
/notes="BAM: BamH1 Q/U"
67477..67649
/notes="Exon in EBNA-1 RNA (Speck and Strominger,1985) and
cDNA clone T4 (Bodscot et al, 1986)"
69410
/notes="BAM: BamH1 U/P"
69684..69930
/notes="5 x 51bp repeats"
70387..70521
/notes="9 x 15bp repeat"
70750
/notes="TATA: CATAAA"
complement(71520..75239)
/notes="BOLFI reading frame, 1 NXT/S analogous to VZV RF

```

```

misc_feature
misc_feature
misc_feature
mRNA
misc_feature
repeat_region
repeat_region
promoter
CDS

```

```

21; NCBI gi: 832979"
/codon_start=1
/translation="MASAMESDSSGGSGADAOQPLAEVDGGLARVTRQLLLSGDDPA
ARLRALMPLELGIFGLDGLAQPVIVROF LNTLTMSGHYPAAVLHNHAYYLRRAASF
SRRSFGLHLEAALDVLAASSLPTTASPAYDDP LDGSRLLAEATRALAAAYRIIEGS
GEVLVAGSPPTATFAVEELVADTYLAWMDAFPREGLSFYAFNAAKTTLGRMLVTVYAE
THRYVAAACQOQPTMAADIKAMAVELVEHSGGAGCGGEGSGGLFHRPESLSVVA
SLPTLARRARVEILGYPAASGGQTPPVAVPVVLAFAARLRLLEPSGALFYDYVEAL
LMDQTYGPDVSIYEAFLAGMAEALAAVQEAAGSRASFSPAALQVATVLLSAGL
NETVAGDYAMMLASVPRSRWRLEATALESISCFALHFEFLIPTASPTSRFAR
VARAAYLRAEACAVDRRARTSGPTAAAPAAATAGVGAADPMDVATPLIRFIVPP
PAAEYEVAGDSGSELLRSYLWYRSYLQWAPAPALPKCPPLPGEQGRROMTAAV
AAAPTDEAVYCRSLRACQTRADAPAVVHSPFEPFAETEQIMPALREVLNSDELKTR
SLAALRWLVSGDSLALPSLUTRARRPLELIYATVMEIYDGNPMPGESPQVAGLRP
RPLDYDETGRVQITPLATEEEEAHVAVKEKSVSSPHYSTDLQTLKSVGEGIQVC
ROAARWALATADTATLRRLLVPLALRESGTADHPIMAHITSEP LRPDLEELNERVEH
ALELCYSLTCALRSSVAVREFDYTFARLQFPALDAERAEAIVRRDARPPVFIIPAR
RIPQCGADTPPLSMDDILYLKGSICKALVDVLDHHPAEPETTP IKTYTPAMDLAPEQ
ITVTPRSPVLAFAARTARVQTHLVLPALTDSDSPSPVQTPPPFRITLPKAKLAILLG
NORNAKSRASRLSPPPHGRWRAVLDSPPSFSSDSFSDQDEGEADLRGVPGGG
CGCAVEEDREPRSDIDTAABAOKVETSCPRRSPTTTPSPSRASGGGSPDRGEAEM
TPPYVLSAAAAAASVRRPRTRGATRRPRPTAEDE"
promoter
/notes="TATA: TATTTAA before BOLFI"
73468
/notes="BAM: BamH1 P/O"
75017
/notes="TATA: TATTTAA BO-R1 late promoter before BORFI,
gives 3.9kb late RNA"
75238..76332
/notes="BORFI late reading frame, 2 NXT/S homologous to VZV
RF20; NCBI gi: 832978"
/codon_start=1
/translation="MKVQSGVDRLRQRRIAGLLPPPARLINISRGSEFTDRVGRIVE
EHAQKLSAAAVHVRAGLAPGEVAVAGGSGGSFSWSGRRPVFGDFLIHASFNN
ACATCTLFQKQSDPFGVDVAFPLSLF ILMHGRCVAAVEAGGGLTRMANLLIYD
SPATLADVPDDEGRIVADREHNFITPVGCLVENIKSTYLNKLTIVVHGPVWSKAIPR
STVKVTPVQAFVLDLAWLSGGAGGGGCVFVGIGLQPCPADARLYVALTYEAGPR
FTFFQSSRGHCQIMNIRIYSPS IMHRYAVVQPLHIEELTFCAVACLGTFSTQDQWR
RSAFNYSGLSDVVEIDSFYSNVSDMEVIL"
complement(75322)
/notes="TATA: TATTTAG before BOLFI"
75819
/notes="TATA: TATTTAG before BOLFI"
75838
/notes="TATA: TATAAAG"
75838
/notes="polyA signal: AATAAA"
complement(76126)
/notes="polyA signal: AATAAA"
76169
/notes="TATA: TACATAT BO-R2 early promoter before BORE2,
gives 2.8kb RNA"
complement(76300)
/notes="polyA signal: AATAAA"
76407..78887
/notes="BORF2 early reading frame, 2 NXT/S. Homology HSV
140K ribonucleotide reductase (Gibson et al, 1984) and RF
19 VZV; NCBI gi: 832980"
/codon_start=1
/translation="MATTSHVEHLLSKLIDELVKANSDPADVLACRLHRIKAES
VTHVVAEYLEVSDKFYDEEFFQMRDELTRVSFAQSPAYERIVSGVSLASLYYD
TYLYVRSRQKESQVHFYMRLAGFCATCTCLYAGLRAALQARPARPTEDEMFVDFYE
HLTSQTVCCSTPFMRFAGVENSTLASCLITLTPDLSLSEMDVTQALYRHLGRYLFQKRG

```

```

misc_feature
promoter
CDS
misc_feature
misc_feature
promoter
misc_feature
misc_feature
promoter
misc_feature
CDS

```

GVGVTGACQCKHISLMRMINSVHNYHCKRPSVSAAYNEPWSHQIFKELETKLP
 ENHERCPGFTGLFVPELFFKLFDRTPMSDWYLFDPKQAGDLERLYGEEFREYRRLV
 TACKCGRVSFKSIFSVINCAVKAGSPFILLKEACNAHFVRDLQCEAMNANICAEV
 LQPSKSVATCNLANICLPCINWPAVLAVRAQADTDQDELLALPRLSVTLPGEGAV
 GDGFSIARLDTQCATVACVACSLQAGSPYYDRSDMSGVLGQGLADVFADLQWQYT
 DPFSSJANKIEFHYVFTALCTSLGLHTRKIPFGKOSKYAGWFHMHDMWACTDLS
 IPREIWSRLSERIVHVDGLPNSQFIALMTSGCAQVTCGSDAFYFFYANASIKVTNKEE
 ALPNSRHWIRLDRALNLVGGVRSCLPEALRQRYLRFQYTAFDYNOEDLLQMSDE
 RAPVDVQSQSHSLFREDDAARASTIANLLVRSYELGLKTIWYVCRIKAADLGMWEC
 KASALSVPREQNERSAEQMPRMEPAQVAGPYDMSKPGCEGCGWCPGCGLEV
 CYRQJLSEDDLLLETGDETERACESQ"

misc_feature

/note="BAM: Bam H1 0/a"

promoter

/notes="TATA: TATAAGT Ba-R1 early promoter before BMRf1,
 gives 3.5kb RNA"

misc_feature

/notes="polyA signal: AATAAA, end of 3.9kb late RNA from
 75017 and 2.8kb early RNA from 76169"

misc_feature

complement(78896)

CDS

/notes="polyA signal: AATAAA"

78900..79808

/notes="BMRf1 early reading frame, homologous to HSV 38K
 ribonucleotide reductase (Gibson et al, 1984) and RF 18
 VZV; NCBI gi: 832981"

/codon_start=1

/translation="MSKLLVDRHGEACLTVEHNRNFWAAHIVLTKDCGLKLINE
 RDLFEYKTLFTFAMAEKLVNFINDELVTFSFESHDIDHYTEQKAMENVHGETYANIL
 NMFLDCGDAAMNAYAEAIMADEALQAKISWLROKVAAVTLPEKILVFLITGEGFIS
 SFYSIALNRVGLMPCIGLANNYSRDELLHTRASLLNSWNTAKADPRATWQELF
 RTAVEVEFTAEFACGCVTLVDRAIKQFLLEATADRLGDTQAPLYGTPTPPKDCPLT
 YMTSIKQNFEEQESSDTMLVDDDL"

promoter

/notes="TATA: TATAACA"

misc_feature

79537

promoter

/notes="BAM: Bam H1 a/M"

79840

/notes="TATA: CATAAAT BM-R1 early promoter before BMRf1,
 gives 2.5kb RNA"

79899..81113

/notes="BMRf1 early reading frame. Early antigen protein
 recognised by R3 monoclonal (Pearson et al 1983; Cho et
 al, 1985a); NCBI gi: 832982"

/codon_start=1

/translation="METQTILRFKTKALAVLSKCYDHAQTHLKGGLVQNLVSNYGG
 PRLAAVANAGTAGLISFEVSDAVAEQNHQSPPEAPAVSRNLAIGRTVCIGKELF
 GSAVEQASLQPKYRPGGSRPEFKLTMYDDKVRKSHHTCAIMPYPASDRINREQ
 MIGOVLNQTASGLQKARQGGGVKLPDNLVVTYTSGEACLTIDVKPLSVGP
 YEAFCTPVAKAQDVCAVEAHVVCVAADSAAALSICRIPAVSVPLIREYSGLIAV
 AGLTSSAGDPLDLVSLFNHASEAAASTASEPEKSPRVQPLGTGGLQQRPHVTSP
 SPSPPPPTPTWESPAPETTPSPATPSHSSNTALERPILAVQLARKRTSSERQKQKH
 PKVKQAENPLI"

promoter

80779

/notes="TATA: TATTAA BM-R2 late promoter before BMRf2"

complement(80782)

misc_feature

/notes="polyA signal: AATAAA"

80832

promoter

/notes="TATA: GATAAAA, possible promoter for 1.4kb late RNA
 encoding BMRf2"

81118..82191

CDS

/notes="BMRf2 early reading frame; NCBI gi: 832983"

/codon_start=1

/translation="MFSCKQHLISGACVFCIGLILASTPTWCFVFNALLSIFSPWQ
 THYVRLGPTACLAVALMTLPKHAHRAVTPAIMLNTASALIFFSLIRVYSTWWSA

PCLFANLPLCLMPRIAIEIVYCPAIHQRFELGELLACTIFALSWSRALEVSVA
 FMSPFIFLAGSGSLAGARRNOIYTGLERRRSIFCARGDHSVASKETILHRCPPDL
 LAISALTVIVCVMIVLHVAEVEFGLSRVLPLELGGMASGGLYGHSSITACVMT
 LCTLSVVVYFLEHTLGLPKTKVTLFISIFYFSVVAALSAMRYKIKKFNCPVLHL
 RVVVMCCFVFTCEYLLVTFIKS"

promoter

81751

/note="TATA: CATAAAT"

misc_feature

82180

/note="polyA signal: AATAAA, end of 3.5kb early RNA from
 78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"

promoter

complement(82311)

/note="TATA: CATAAAT"

repeat_region

82319..82461

/note="2x71bp repeats"

CDS

complement(82746..84122)

/note="BMLF1 early reading frame. Diffuse early antigen
 (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
 of HSV (BSIF2 + BMLF1) is also called ER2
 (Chevallier-Greco et al, 1986). General transactivator of
 transcription (Lieberman et al, 1986)."

misc_feature

complement(82747)

repeat_region

83640..83729

/note="polyA signal: AATAAA"

misc_feature

complement(84122)

/note="ACCEPT: CTCCTCTCCAG acceptor in spliced form of
 BMLF1 RNA"

misc_feature

complement(84227)

/note="DONOR: CAGCTAAGA donor in spliced form of BMLF1
 RNA"

CDS

complement(84229..84288)

/note="BSIF2 early reading frame in 5' exon of spliced RNA

misc_feature

encoding BMLF1; NCBI gi: 832984"

/codon_start=1

/translation="WPSQRLSRTSSISSNEDPA"

CDS

84233

/note="BAM: Bam H1 M/S"

complement(84257..86881)

/note="BSIF1 reading frame, homologous to RF 6 VZV; NCBI
 gi: 832985"

/codon_start=1

/translation="MSAPVVKIIVASNTDIAEILDAILSRDEGFRLECLCHNASP
 LHHVAGSLVELQLHLPKRLTSQSCGLVTLHLPAEAPFFLRGLTPLADRLSTYL
 DRAGALRSJLTPVELLTLISAKQPOGARGVAMLRPKIVGCLIRYRVNISARMFIS
 TFGSHEAQFVLTAAVYFWGIPCTIETLAHLELFTSFGSLAAVTSIAELGEVGS
 SWAQCEAFANFAHEKLRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHQCF
 NEGTFKYSQJLTSMTGCLPSGGVYLPDLLDRGFAEHMRTYFTRETYLAEHVQOOL
 KTRMEPPAPYTWDPDDGLMARWAGLSVDARELVELARHWADEGCTPTPTLQGFCL
 LAQATCRGQWNPKEQFLPTTVLRVQRLPFLCHFAHRYFVMTAADPFSHIAEV
 STPNCLPDTLCTALSYTPVYYSQNSLSEQLFVSRHEYFNRLPVCNLVLDLDKI
 KGAPSLSEETDLCRTVRREVLRILMRIGPVRAPVYFFKACPPADPNMEDVLPF
 CICTCKLGRFVITPLRCHAIYCTSAVQGVSVLQKLMGLTACTLRMRHRIKEIGAPL
 FDSYGHACRILPHTYTKVDRGGLSRQLRFVCHPEEDKHYSVRNALNQLLHH
 SLRVGWPAKPTFYIADDDQYLIORTRETLPTPTVENVCAMTEGLGLDILVAVSSC
 IMPSLMSTLATVAPDEKPEQFLHVTFEQTPNLVQVCHARGNFACLRHTRHASKNV
 RFVLVLYTTSQAITVTMSQCFAGRCGANOPTAHFSISVPASRIINRAESQSSTTSQ
 LARRDRQDGSFETLPN"

promoter

complement(84356)

/note="TATA: CATAAAT before BSIF2 and BMLF1. Two RNAs
 start here; one is spliced and the other is unspliced,
 both traverse BMLF1."

promoter

86882

/note="TATA: TATTAA BS-R1 late promoter before BSRf1"
 86924..87580

CDS

```
misc_feature
misc_feature
misc_feature
CDS
misc_feature
promoter
promoter
promoter
CDS
promoter
CDS
misc_feature
misc_feature
CDS
```

```

/notes="BSR1 reading frame; NCBI gi: 832986"
/codon_start=1
/translation="MAFYLPDMSCCGLWLFGRPNRYVSOLPEETFECPDRWRAEID
LG1PCVQVQGLLRNQWTCMSLRVYLLAVOANSITDHLKRPDAVRVPSRCRGVRAQ
VAKLEAVRSVIMTWISLAVSGTEMDENGKALKLDKQAGDSIALMEMEKVATALKWDE
TGAQAQESAVSVSVTAPASAPFINSAPFINSAPFINSAPFINSAPFINSAPFINSAPF
"
complement(87134)
/notes="polyA signal: AATAAA"
87599
/notes="polyA signal: AATAAA"
complement(87613)
/notes="polyA signal: AATAAA, end 1.0kb early RNA from
BLI3"
complement(87638..88474)
/notes="BLIF3 early reading frame (BLIF2 in Baer et al,
1984). Homologous to RF 8 VZV and dUTPase HSV. NCBI gi:
832987"
/codon_start=1
/translation="MEACPHIRYAFQNDKLLIQQASVGRLLTVNKTILLRPMKTTTV
DUGIYARPEEGHGLMIGSTRPVTSHVGIIDPGYTGELRLIQQRNYNSTLRSEL
KGLIAFYATYPMQEDKGPINHPQYDGVGLDYSLPKDLALFPHOTVSVTLVTPPS
IPHREITFCGSLAQGLLVKPCRRRCGVSVLNFSDQTVFLNKYRRCQLVYLH
KHLTFSYPSHDAGVLGRSLRFWASCTFEFVSLAMGDSGLSEALGROGRGFGSS
CQ"
87650
/notes="BAM: Bam H1 S/L"
88507
/notes="TATA: TATATAT BL-R1 late promoter before BLRF1,
gives 1.0kb late RNA"
88511
/notes="TATA: TATAAGA"
complement(88514)
/notes="TATA: TATATAT BL-L3 early promoter before BLIF3,
gives 1.0kb early RNA"
88547..88855
/notes="BLRF1 late reading frame; NCBI gi: 832988"
/codon_start=1
/translation="MGKVLKPKFAKAVPLFLAATWLLTGVLPAGASSPTNAAASIT
EAQDQFYSYTONADTFSSLSFASIWALLTLVLVLIASAIYLMYCFNKFWNTLLTD
"
88863
/notes="TATA: TATTATA BL-R2 late promoter before BLRF2,
gives 0.6kb late RNA"
88925..89413
/notes="BLRF2 late reading frame, 2 NXS/T; NCBI gi:
832989"
/codon_start=1
/translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLIESENKALKQOVLRCG
ACASSTVSPAPVPPPEPTARQREWNITQATGLASQAMKKIEDKVRKSVGVGTNRN
EMENILQNLTLRIQVSLMGAGQPSPECTPRESDNPATRRARSRSRGREAKYQVI
SD"
89412
/notes="polyA signal: AATAAA, end of 1.0kb and 0.6kb late
RNAs"
complement(89425)
/notes="polyA signal: AATAAA, end of 0.7kb early, 2.2kb
late and 2.8kb late RNA"
complement(89430..92153)
/notes="BLIFla, late reading frame, gp350 membrane antigen,
36 NXT/S (Hummel et al, 1984; Biggin et al, 1984; Beisel
et al, 1985); NCBI gi: 832992"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFFPYTCNVCTA
```

```

DNNVTINFVGKKHQHLDLDFGQLTPHTKAVQPRGAFGGSNAIYLFLELIGAGEL
ALTMRSKKLP INVTTGEEQVSVLESVDVYFQDFVTGMCHAEQMPVYLIPETVPYI
KQDCNCSNITAVVRAOGLDVLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDEGEI
SQVLCKNKENITCSGESHVSPGGLITSISPVATPIGTGYAYSRLTRPPVSRFLG
NNSILYVYSGNGKASGDYCIQSNIVFDEIPASQDMPNTTDDITYGDNATYSVP
MYTSEDASNPNVTATWAMPNNTEFDCKKWTLSGTPSCCENISGAFASNRTEDIT
VSLGTAPKTLIITATNATTTTHKVFISKAPESTTTSLNTTGFADPNNTTGLPS
STHVPNLNLTAPASTGPTVSTADVTSPAGTTSASPVTPSPWDNGTESKADMTS
STSPVTPPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTP
PNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAV
ATNHTLGTSPTSPVVTQPKNATSAVTTGQHNITSSSSMSLRPSNPETLSPSTSD
NTHSMPLLTSAHPGTGENTITQVTPASTISHTVSTSSPAPRPGTTSQASGFGNSSTST
KQCEVNTKGTTPPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAV
PTTDYGGDSTPRPRYNATYLPSTSKLRPWTFSPVTTAAQATVPVPTSPQRF
SNLSMLVQWASLAVILLLLVLVMDACAFRRNLSTSHYTTTPPYDDAETVY"
complement(89430..92153)
/notes="BLIF1b, late reading frame gp220 membrane antigen,
spliced form of BLIF1a (Hummel et al, 1984; Biggin et al,
1984; Beisel et al, 1985); NCBI gi: 832991"
/codon_start=1
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFFPYTCNVCTA
DNNVTINFVGKKHQHLDLDFGQLTPHTKAVQPRGAFGGSNAIYLFLELIGAGEL
ALTMRSKKLP INVTTGEEQVSVLESVDVYFQDFVTGMCHAEQMPVYLIPETVPYI
KQDCNCSNITAVVRAOGLDVLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDEGEI
SQVLCKNKENITCSGESHVSPGGLITSISPVATPIGTGYAYSRLTRPPVSRFLG
NNSILYVYSGNGKASGDYCIQSNIVFDEIPASQDMPNTTDDITYGDNATYSVP
MYTSEDASNPNVTATWAMPNNTEFDCKKWTLSGTPSCCENISGAFASNRTEDIT
VSLGTAPKTLIITATNATTTTHKVFISKAPESTTTSLNTTGFADPNNTTGLPS
STHVPNLNLTAPASTGPTVSTADVTSPAGTTSASPVTPSPWDNGTESKADMTS
STSPVTPPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTP
PNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAV
ATNHTLGTSPTSPVVTQPKNATSAVTTGQHNITSSSSMSLRPSNPETLSPSTSD
NTHSMPLLTSAHPGTGENTITQVTPASTISHTVSTSSPAPRPGTTSQASGFGNSSTST
KQCEVNTKGTTPPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAVTTPNATSPPTAV
PTTDYGGDSTPRPRYNATYLPSTSKLRPWTFSPVTTAAQATVPVPTSPQRF
SNLSMLVQWASLAVILLLLVLVMDACAFRRNLSTSHYTTTPPYDDAETVY"
complement(89434)
/notes="TATA: TATAAG"
complement(89567..90013)
/notes="BLIF2 early reading frame (BLIF3 in Baer et al,
1984); NCBI gi: 832990"
/codon_start=1
/translation="MCPVVRQHPAQAPPAKQALETVHPQNRGLMSPKARPPKMQR
RRPPVAKRRFRSPQOQVERILPPVESTPDQMEPQGVQSPQITAVIQLRQDRDTH
RPPIYLPALLANGCPAGLIRAHRLPQPKPCQSRQRPSPDSQTSPC"
complement(90051)
/notes="TATA: TATAACA BL-L2 early promoter before BLIF2,
gives 0.7kb early RNA"
complement(90062..90652)
/notes="intervening sequence in gp220 gene"
90177..90639
/notes="21 copies of 21bp approximate repeat"
complement(92192)
/notes="TATA: TATATAA BL-L1 late promoter before BLIF1a,b.
Gives 2.8 and 2.2kb late RNAs"
92238..92581
/notes="Exon in Bodescot et al (1986) RNA (spliced from
20763 to 92670)"
92243..92581
/notes="BLRF3 reading frame; NCBI gi: 832993"
/codon_start=1
/translation="MDKORCGPALDONMEEVEPSTSVQEQVSAGDWENVLIELSDS
SSEKAEADHLEPAQKGTREKRVDDHAGGAPARMLPQPQDLPGREAIURRFPDLRL
TLIQAIGAAAT"
```

```
CDS
92670..95162
/notes="BERF1 frame, homology with BERF2b and BERF4. A
fusion of BLRF3 with BERF1 encodes EBNA-3a, latent cycle
gene. (Hennessy et al, 1986, Joab et al, 1987)"
92670..95248
/notes="Exon in (Bodescot et al, 1986) RNA from 92581, to
3' end"
92703
/notes="BAM: Bam HI L/E"
/notes="complement(93161)"
/notes="TATA: CATAAAT"
93479
/notes="TATA: TATAAGA"
/notes="TATA: TATAAGA"
complement(93482)
94208..94277
/notes="TATA: TATAAAT"
94208..94277
/notes="repeat type A"
94281..94306
/notes="repeat type B"
94307..94381
/notes="repeat type C"
94386..94411
/notes="repeat type B"
94412..94489
/notes="repeat type C"
94490..94560
/notes="repeat type A"
94571..94648
/notes="repeat type C"
94649..94719
/notes="repeat type A"
94896..94982
/notes="repeat type D"
94983..95069
/notes="repeat type D"
95221
/notes="polyA signal: AATAAA"
complement(95272)
/notes="polyA signal: AATAAA"
95353..95724
/notes="BERF2a reading frame; NCBI gi: 832994"
/codon_start=1
/translation="MKAWLSRAQQADAGCAGSEDPDYGDQGNVTQVGSEPISEI
GPFELSAASEDDPQSGPVEENLDAAREEEEPHEQHGDDPLDVHTRQPRFYDVP
TOAPVIQLVHAVYDSMLVRGT"
95725..98244
/notes="BERF2b frame, homology with BERF1 and BERF4. BERF2a
and BERF2b are spliced together to make EBNA3B (EBNA4A)
latent protein."
complement(95819)
/notes="polyA signal: AATAAA"
complement(95853)
/notes="TATA: TATAAAT"
complement(96276)
/notes="polyA signal: AATAAA"
97522..97698
/notes="3x60bp repeat"
98323..98769
/notes="BERF3 reading frame; NCBI gi: 832995"
/codon_start=1
/translation="WTILDEVEILHIFRPTMESFEQGDQRSPDNRGDNVQTGGEH
DQDPGCPSSGASERLPEESYRQQPWPQSGSDENCMQRIIRRRRRRAALSCH
LLQEDNVFWPLPHDITFTYARNIRDAACRAVKVSMPLTGMGAI"
98364..98730
```

mRNA

```
misc_feature
/notes="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
98731
/notes="DONOR: AAGGTGAGT donor"
98805..99050
/notes="Exon in T4 cDNA (Bodescot et al 1986) . 99050 is not
the end of the RNA."
98805..101420
/notes="BERF4 frame, homology with BERF1 and BERF2b. BERF3
and BERF4 are spliced together to make the EBNA3C (EBNA
4B) latent protein."
99126..102118
/notes="DEL: Deletion in Raji"
99443
/notes="TATA: CATAAAA"
100104
/notes="DONOR: ACCGTGAGT possible donor before repeat."
100122..100304
/notes="10 x 15bp repeat"
100528
/notes="DONOR: CTGCTAAGG possible donor"
100613
/notes="BAM: Bam HI E/e1"
100665..100781
/notes="3x39bp repeat"
complement(100860)
/notes="TATA: TATAACA"
100919
/notes="BAM: Bam HI e1/e2"
101426
/notes="BAM: Bam HI e2/e3"
complement(101445..102116)
/notes="B2LF2 reading frame 3x NXT/S. 2.5kb late RNA
traverses B2LF2, ends unknown. NCBI gi: 832996"
/codon_start=1
/translation="WVSFKQVRVPLFTAIALIVILLVFLPVRVGGRRVAAAALTW
VPKPNVEWVPDPVPNNFNKTAQEKVKLPHWTFPLHTFQVPQNTKANCYTC
NTRYETSYKGCIFYTKKKHMGCAQCAELPYCTFYGTPDILPVTNRINAIE
SLWGVYRVGEGNWTSLDGTFRVYQIFGSHCTVTSKFTVPVSHHESGFLKPCLCVS
QBSNS"
101690
/notes="TATA: CATAAAA"
101765
/notes="polyA signal: AATAAA"
complement(101786)
/notes="TATA: TATAAG"
101947
/notes="BAM: Bam HI e3/2"
complement(102098)
/notes="DONOR: CAGCTGAGG possible donor"
complement(102126..102341)
/notes="3' terminal exon of 0.9kb and 2.8kb early RNAs"
102153
/notes="TATA: TATTAAT"
complement(102156)
/notes="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs
encoding B2LF1 and BRFL1"
complement(102160)
/notes="TATA: TATTAAT"
complement(102213..103155)
/notes="B2LF1 reading frame, modified from Baer et al,
1984. Has two splices within frame. 2xNXT/S. Immediate
early gene which disrupts latency (Countryman and Miller,
1985), called EB1 by Chevallier-Greco et al, 1986 and
ZEBRA by Miller."
```

CDS

```

promoter      complement(102380)
              /note="TATA: CATAAAT"
promoter      102415
              /note="TATA:TATATAC"
promoter      complement(102420)
              /note="TATA: TATATAC"
mRNA          complement(102426..102530)
              /note="Exon of 0.9kb and 2.8kb early RNAs"
misc_feature  complement(102504)
              /note="polyA signal: AATAAA, apparently not functional"
repeat_region 102581..102652
              /note="semi-repetitive sequence, homologous to human c-fos
              3' sequence"
mRNA          complement(102655..103194)
              /note="First exon of 0.9kb early RNA encoding B2LF1"
misc_feature  complement(102918)
promoter      /note="splice acceptor used in RZ fusion gene (Sargeant)"
              complement(103231)
              /note="TATA: TTTAAA of B2L1 immediate early promoter gives
              0.9kb RNA"
misc_feature  complement(103256..103311)
              /note="Upstream of B2L1, homology to 106243 to 106188"
CDS           complement(103366..105183)
              /note="BRLF1 reading frame, (immediate?) early gene, acts
              as transcription activator. NCBI gi: 832998"
              /codon_start=1
              /translation="MRPKKGLEDLRLTPEIKKQLGSLVSDYCNVINKETAGSVEI
              TLRSYKICAKINAKAGREMGGLATLINCNFWAILNNRVRRRAENAGNACSITA
              CPTVMRYVLDHLVTDTRFFTOAPSNRVMIPATICTAMYKLIKHSRVRAVTVSKVLGV
              DRAAIMAGSKQVHEHLNRMKEGILLSKFAKCKWFTTPVLEEMQTNWSKGTGHLT
              DDVKVRALIKITLPRASYSHAGORSYSGVLPACLLSTKSKAVETPILVSGADRME
              ELMGNDGASHTARYSESGQFHAFTDELESIPSTWPKPCQASADCGDSSSSDS
              GNSDTQSEREARAEARLRAKSRRTSRNRGQTPCPSNAEPQPHIAVHQESD
              ERIFPHSKPTFLPVRKKGLRSGRMFLPKPAGSAISDVFEQREVQPKRIRP
              FHPGPMANRPLPASLAPTPGTGPHVPSLTPAPVQPLDPAVATPEASHLLEDP
              DEETSQAKVALREMDVTIPOKEAAICQQMDLSHPPPGHLDLITTTLESHTDLNL
              DSPLPENELTDTFINDDECLHAMHISTGLSIFDTSLF"
              complement(103453..103462)
              /note="TAATGAAATC sequence"
              103741
              /note="BAM: Bam H1 2/q"
              103816
              /note="BAM: Bam H1 g/R"
              complement(104926..105185)
              /note="exon in RZ fusion gene (Sargeant)"
mRNA          complement(104927..104989)
              /note="BRLF2 poss. small 5' exon"
              105016
promoter      /note="TATA: TATAAAT before BRRF1, possible promoter for
              1.1 kb early RNA encoding BRRF1"
              complement(105019)
promoter      /note="TATA: TATAAAT before BRLF2"
              105182..106114
CDS           /note="BRRF1 early reading frame; NCBI gi: 832997"
              /codon_start=1
              /translation="MASSNRGNARPLKSFLEHLYIKHPEYGVNHLNTIGVDCDLP
              PSHPLTAQRGELIARVLAQVQKHLEDTIVPKLIKLAFLLELLSYSPKQDRDI
              AEYLDHILKTNRLDLDRLWALIRKLRDRHSAVNVLMGSDYTAVSLOYDGISIG
              MRKVINDVCRSGVSNPSTWATNLSHQLLMASCPSEECWAGRFQNVLLWTVALCK
              FRCIYNYIQGSIATPISQLLHLEIKALCSWITSQDMRLFQHSRPLLTLMESVAANQ
              EVTDATLTPDCAEVIDLLKTKHVLNCSAMQYK"
              complement(105185)
              /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
              BRLF1 and RZ fusion gene (Sargeant)"

```

```

promoter      105213
              /note="TATA: CATAAAA"
misc_feature  106110
              /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
              encoding BRRF1"
              complement(106125)
misc_feature  /note="DONOR: CAGGTAAGA possible donor"
              complement(106188..106243)
              /note="Homology to upstream region of B2L1"
promoter      complement(106213)
              /note="TATA: CATAAAA"
              106243
promoter      /note="TATA: TATAAAA before BRRF2, possible promoter for
              1.8 kb RNA encoding BRRF2"
              106302..107915
CDS           /note="BRRF2 reading frame; NCBI gi: 832999"
              /codon_start=1
              /translation="MSQQRGSVILVPEHLAAGALTKLMSDFITGQDVTLSGQNIKVI
              RDAINOTPGGGDVALISLSIFALMNALPTSGRSSHODLIPAAVOALTAAHNLCLGVIP
              GETSKHPTESLIRAITVGLQKLWDSGCCPECLQACIKGLKAIKPLGYEIPRIIPHTK
              QCSVNLINMLVHLKVALRGVQLAYDARVLTPDFEIPDLDDSDAVFARTLLAALFH
              LNWFTILKYDITQDSMLKQALSGHWSATGNPLPAAPETLRDYLEAFRNSDNHFYLP
              TTGPLNTQFPPELLGRVWVVDSSILCAASHVQDVITHGVGAGVPRFRSALPAPSR
              PQTCSQLTSRGNESSRRNIQPGGTSPVPPVCPVLSLTASGAKQNRGSGSLHLAK
              PEETSPAVPVPCTIASPASRSKHQCVGTGSOAPSFSSVAPVASLSGLDEEEEGS
              RESPSLSKKGDEEFEEAWLEADANLEDVQREFSGLRVIGDEDEGSGCEFSDDL
              SUSDHEGEGGAGVGGGRSLHSLSYSVW"
              complement(106385)
              /note="TATA: CATAAAA"
              complement(106973)
              /note="polyA signal: AATAAA"
promoter      complement(107124)
              /note="TATA: CATAAAA"
              107457
misc_feature  /note="BAM: Bam H1 R/f"
              107565
              /note="BAM: Bam H1 f/K"
              107914
misc_feature  /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
              BRRF2"
              107942
              /note="ACCEPT: splice acceptor for EBNA-1 RNA (from
              98730)"
              107950..109875
              /note="BRRF1 encodes EBNA-1 protein, latent cycle gene.
              NCBI gi: 833000"
              /codon_start=1
              /translation="MSDEGFGTGCNGIGEKGDTSPEGSGSGPQRRGGDNHGRGR
              RRGRRGGRCFAPGGSGSGFRHRDGVRRPQKRRPSCIGCKTGCTGACAGAGAGG
              AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
              SPRRPPGRRPFHFVCEADYVEYHOEGGDCPEVPPCAIQGQADDPCEGFTGP
              RQDGGRRKGFCKHRQCGGSNPFENIAEGLRALLARSHVERTTDECTWAGVF
              VTGSKSTSLYNLRGTALAIQCRLLPLSLRPLFMAPGQPGQPCPLRESIVCYFWFL
              QTHIFAEVVKDAIKDLVMTKPACTCNIRVTVCSFDDGVDLPHPFPVMEGAAGDGG
              DQDEGDEGEDEGEQ"
              108217..108924
              /note="EBNA triplet repeat GGA, GCA, GGC."
              109856
              /note="DONOR: AGCGTCAGS possible donor at end BKRF1"
              109905
              /note="DONOR: AGCGTCAGS possible donor at end BKRF1"

```

```
misc_feature /note="TATA: TATTAA before BKR2, possible start for
2.3kb late RNA"
misc_feature /note="polyA signal: AATAAA"
CDS /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
109958..110371
/note="BKR2 reading frame; NCBI gi: 833001"
/codon_start=1
/misc_feature /translation="MRACVFLAICLVTFVLPWGNWAYCCCHVTQLRAQHLLALEN
ISDIYLSNQCOCGFSIASINSRPNQSLVSRKANGLVNVSFFISILKSSSALTG
HLRELITLTLYGSFSEDLFGANLRYAHRGG"
110271
misc_feature /note="DONOR: TCGGTGAGT possible donor at end of BKR2"
110275..111117
CDS /note="BKR3 reading frame, homologous to RF 59 VZV"
111098
misc_feature /note="DONOR: TCGGTGAGA possible donor at end BKR3"
111107..111784
CDS /note="BKR4 reading frame, contains complex repetitive
sequence"
111272
misc_feature /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
BKR4"
111719
misc_feature /note="polyA signal: AATAAA"
111787
CDS /note="polyA signal: AATAAA : currently unknown which is
3' end of the 2.3kb late and 1.1kb early RNAs"
/note="BBLF4 early reading frame, very good homology to
RF55 VZV; NCBI gi: 833003"
/codon_start=1
/translation="MAEPRAPALSTFTMLNWTSDASVRIRVRRIGTLARRVQQLP
DMETSPFDELPSELPFPFSAYVITGTAGAKSTSVSLHPTMCLVTGATTVAAQ
NLSTQIRACPTVYSAGFKSRHINMTORVSSHRSRSTDAALELQRRDLAKWPLVSD
IAEFRRTRKGLSGVSGPAFEVLEDRMHQQLMTNVIWVDEAGTSLWHILTAWFC
YWFNNALRPTLYRRRCIPVCVGSPTQDAFQSSFSETQVNKRDCMDILFTLVG
NPAATYVDVARWALF INKRCVDQVGFCHLMKTLEYGLELSPDILAYVDRVVPRAA
IMDPAQYGVNTRFLSHAENVKTFLLTHATLKTAGAGRAAGTGGGGGVTFCTPVE
CEYFLDPLAQTKTLVGLPGLTAHTWLQKNYARLGNYSQADQMDVPVTEQDEERVK
TNNVTVKHSVSWNCKTKSICGCTGTFGDFMDTLEADSFEAHGHEQPEVYSLA
RLIYGIVAFSHGHSLCENGCEYVAGLGNVPLGRTWDEVTAGMELGELPLEVAMDG
ERSPAAVYARVLAPPANASAPLCSLNIYNDLRAVRCQLDVAVRGREFRDLPC
TFNNMLIRDNLEFSTDEPLHRLGLDYASTTENYLLGLVNVFFGIRGKQOPDAG
SSRMPRLWDEAGFVCCLEHNKLYETIEDKSLNLCISIRDYGISSKLMTAKAQQ
LSLNKVAICFGSHRNITKFGHVYALLSRASHNCVMDRNP LSEMITCEGNPASCYVD
ALKNSRALLY"
misc_feature complement(111830)
promoter /note="polyA signal: AATAAA"
112471
promoter /note="TATA: TATATAT"
complement(112476)
misc_feature /note="TATA: TATATAA"
112620
promoter /note="BAM: Bam HI K/B"
113876
promoter /note="TATA: TATTAT before BBRF1"
complement(1113885)
CDS /note="TATA: CATAAAT"
114204..116045
/note="BBRF1 late reading frame, homologous to RF 54 VZV;
NCBI gi: 833002"
/codon_start=1
```

```
/translation="MFNNVDESAGALGSSAIPVHPPTASVRLFEILQKAYVQGG
TIYANLRNPGVFSRQVFTHLFRAISHCTYDDVLHDMNKEACIQKRWPSDDSCASRF
RESTFESKSTTKMLTVRLDILTTNIVRVLHRSRVSLSVERYVDWICATGMPAVKPKITQ
ELHSKINSUDRCVCRELCHERTIRSIGTELYEATKEIIESINSITFIPQFTEVTIYL
PRSEYVAYYCCRRIRLHVLPFAIFACTVFDSPVRLYQIFMCTYRTEHAKICQL
LNTAPLKAIVHGGDMYKDI LAHLEONSQKDPKKELLALLKLENKTTISGVTDVY
EEFITDASNVLDRNRILFCGPGETAAQGLKKKYSNTVVKCLTDQINEQDQINGLEKE
RELYLKKIRMSQGLQASLGP GGNPAASAPAAVAEASVDILTGSTASAEKLFNS
PASIGARVSGHNEISINFSVQIYPPSRGWTOKDITELMESELFNTFKLTPVVDNQQG
RLYVYSSDTSISILIGPFTIYVAELSPVELVTDVYATIGLIVEIIDLRYSSRLAIYIE
DLGRYICPASATGGGHHIRQAPSARGTEPDHAKSKPAROPPPGAGS"
115843..116781
CDS /note="BBRF2 late reading frame, homologous to RF 53 VZV"
/note="complement(116683)
promoter /note="TATA: GATAAAA"
misc_feature complement(116696)
CDS /note="polyA signal: AATAAA"
complement(116784..117386)
/note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
contains a consensus nucleotide binding site"
116785
misc_feature /note="polyA signal: AATAAA"
intron complement(117386..117515)
/note="intron spliced out in RNA linking BBLF2 and BBLF3"
CDS complement(117515..119080)
/note="BBLF2 early reading frame, spliced to BBLF3"
118981
promoter /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
119067
promoter /note="TATA: TTTAAAA BBR2 late promoter ?"
119098
promoter /note="TATA: TATTAA BBR3 late promoter before BBRF3"
119108
misc_feature /note="DONOR: AAGTCGAT possible donor"
119137..120354
CDS /note="BBRF3 late reading frame; NCBI gi: 833004"
/codon_start=1
/translation="MKSSKNDTFYRTWVTKTLVTVFMVNSAVPITAMFPNLGYPC
YFNALVDGALNTYNLAHLPTLYLEPPENEVYITLVFIADCVAFIYACEVAL
IKARKVSGLTDLSAWSVSGSTVIFLAIKLMSIQVF IQVLSYKHVFLSAFVFLH
FIASVLHACACVTFRESVWYVKAQDMSIPQDTFLAMVVFYKRPVVTNLYLGLALETL
VTSLSVFLALGNSFYEWGDMVLGAVNLELILP IFWYILTEVWLASFURHNFGEYCGM
FIASILLILPVRVYANFVSAKLHTTVAIVAIIPILCSVAMLIRICRIFKSRQGTD
YVPVSETVELESEPRPRSRTPSPGRNRRSSSTSSSSSRSTRQRVPVSTQALVSV
LPMVTDSEETFP"
12060
misc_feature /note="ACCEPT: ATCTTCTCCAGT possible acceptor"
120358
misc_feature /note="polyA signal: AATAAA"
CDS complement(120747..120974)
/note="BBLF1 late reading frame, possibly homologous to RF
49 VZV; NCBI gi: 833005"
/codon_start=1
/translation="MGALWSLCRRVNSIGDVGGINLYNDYEEFNLETTKLIAAEE
GRACETNEGLEVEDSENDLELFLPNKPN"
misc_feature complement(120764)
CDS /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
early, 3.0kb early RNAs"
complement(120929..122341)
/note="BGLF5 early reading frame, homologous to RF 48 VZV
and alkaline exonuclease of HSV; NCBI gi: 833006"
/codon_start=1
/translation="MADVDELEDPMEEWTSYTFARFLRSPETAEFVRNLRPPQPMAM
REVVYLCLKQIQEFGSETGFCDFVSSLIQVENDSKDGSILKSIYWGIGLEATDQRVL
```

CSVSNTRGSGENLMDILANGIISSSKLLSTKNGPKTKPEPAPISTNHHFCGPA
 FGLRCDTVDKIVCKLICDASNRQFGFMISPTDGGIFGVLIDLCYNVESQDFILFT
 DRSCEIYKREYLFESKEDPIYPSTALYKRPCKRSFIREINSIRPTVEVPDG
 RLPSEGYILKQDEANWIKDVRKKLPGCHDLVADSIAANRGVESMLYMTDPSEAG
 RIGCKDVPVNIIFINPRNRYFYQVLLQYKIVGDIYVRHSCKGKGRDCSPRNIIVTAF
 RKSPIDPATCTGLSDLLDASVEIPVAVLTPVPLPDSVIRKTLSTLAAGSMKAYDN
 TFDATWPSPGLEADDESTP"

promoter
 complement (121331)
 /notes="TATA: TATTAAA BBL1 late promoter before BBLF1"

promoter
 121669
 /note="TATA: CATAAAT"

promoter
 121697
 /notes="TATA: TATAAG"

promoter
 121772
 /notes="TATA: CATAAAG"

misc_feature
 122313
 /note="BAM: Bam HI B/C"

CDS
 complement (122328..123692)
 /notes="BGLF4 early reading frame, homologous to RF 47 VZV"

misc_feature
 complement (123506)
 /notes="DONOR: AAGGTGACT possible donor"

CDS
 complement (123941..124939)
 /notes="BGLF3 reading frame; NCBI gi: 833008"

/codon_start=1
 /translation="MFNAVKADMPDDPMIARRYQCILEALEACQDTPQFKLWETPL
 KSFLLVSNILPQDNREPMHARSGRVAEDDYDFSSIALLELPLINPLRPEEQFGQGW
 SSMEPSPEMGCLCFEVEDGIMRIALANKDEVIGQALQIILANSQWTSINVEDP
 LPMWALYCPRSRCHCEERHCYAAAKRGKPIILPTAVTFCANIEAFLAHLTCVYA
 LYLDVRMKGEDIAPPDFVSLRANKAKOICLLPOEPFCITRVCLLCLLHKONLNAQYK
 RPVDYDCLILTGEAERYWDAVGNREASTGTGTVLYTYDLSIVADMVTEDE"

promoter
 complement (124117)
 /notes="TATA: TATAAAA"

misc_feature
 complement (124219)
 /notes="polyA signal: AATAAA"

CDS
 124938..125915
 /notes="BGLF1 reading frame, homologous to RF 45 VZV and
 spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
 Northern blots in BDRF1 detect 2.7, 2.6, 2.1kb late and
 1.9kb early RNAs. 2.6, 2.1kb RNAs very weak. NCBI gi:
 833007"

/codon_start=1
 /translation="MLYASQRGLTENRNALQDSTTGCCLGAEPTSMYTGAKSDR
 WAMPVGTIHASNLCPMLRAYCHGPRPVVASDESIPMGASALHTPVQVQCL
 LPELRTLQRLPPNLEDEALTEFTKSVSSARAILEDPNFLEMREFVTSIASFLSG
 QYKHKPABLEAFQKQVLLHSFYLIISKSLEITDPMFDIFQSAPGLEEMTEKLIHFK
 QKASVFLIPRHHGKTWVVAIISILISLNSVQICVVAHQKHVASAVTEIIDTLTKS
 FDSKRVENKETSTITFRHSKISSTVMCATCNKNRVDVSVLGNCA"

promoter
 complement (125113)
 /notes="TATA: TATAAAT before BGLF3"

misc_feature
 complement (125484)
 /notes="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
 late, 3.0kb late and 3.7kb early RNAs"

CDS
 complement (125863..126873)
 /notes="BGLF2 late reading frame, poor homology to RF 44
 VZV; NCBI gi: 833009"

/codon_start=1
 /translation="MASAANSREQLRKFINKECILMVLSDASTPMQKVVYTTAVASV
 VYVQIAGPKPYNNVTLVIKPKKPYVTVYINGTLATVARPEVLFKAVQGPHSIL
 LMYGVSDAVCAVPEIRGNVPWCTDLTTAVHTFISTAVKVEELQDITPELIP
 LGRGAWFAEGALYMFVNMMDMLCCPNPPTSLTHFINLLTRCDNGECVTCYAGA
 HVNILRGWTDSPGTSCTCPLCPCTALNDNDYVPIIGHRALGLMKRPEDAPFVGL
 RFPNPKMHPDMSRVLQGVLANGKEVCTAQPTWLLRFSDLYSRAMLNQCQVLRQVHL
 SY"

promoter
 126277

CDS
 /note="TATA: GATAAAA"
 complement (126851..128374)
 /note="BGLF1 late reading frame; NCBI gi: 833010"

/codon_start=1
 /translation="MDVHIDNQVLSGLGTPLLVHLFVDPVTMAELCPNRVNCCEGAWC
 OTLFSDRITGLTRVCVFAARGMLPGRPSHRGTFTSPVYCGEGLPCLNPPHVAALRE
 YDEGGLGEIQIYIISLFEFAGKALTDGHLIREASGVQESAAAMQPIIDPGRPAGAG
 IEMFVAAQVHPKTYDIAKOILILETOENRGEQRLGHAGSPALCLGLRLRAGAEVK
 AAATSVSKHHPALENPNIRSGAGCGGAGCTGTVGVGSGALSRSVPVFSKTRR
 AIRESBALVRGTAHIFSPHALYVTVYTPSLNAQGRLLHMTAVTHASPDIDLAEVILGA
 PEREFRLISVALRISAFREKLQAMQTAQGEIPWIPITSYISRYKNSDLIREAFT
 VQTRVSWESCKATISNAPKTPDACLWIDSHPLYEEGASAMGKVIDSRPPGGILVGA
 SQLVALGTGCHVCVHLATTSQQAFLVLPGGFEVVKQLALTPEERGVYLARHGIRREQ"

promoter
 complement (126929)
 /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
 gives 1.6kb late RNA"

promoter
 complement (127237)
 /note="TATA: TATAAAA, potential promoter for 1.8kb late
 RNA"

misc_feature
 128029
 /note="polyA signal: AATAAA"

CDS
 complement (128344..129021)
 /note="BGLF4 early reading frame; NCBI gi: 833011"

/codon_start=1
 /translation="MSDQGRLSIPRGEGTDEPNRHLCYSKLEFHLPIPESMASVF
 ACKWCGEYHVDCSSSECTLIEHGVCAITGNYPGHPQALRPWTEIRQDTQDRD
 KNEPQVQGLVITVYNNHLYHFINENYISGVSEALFDOEGALRHPALPALSFFVPCCL
 MLFRGASGEKVDVLSLYIHVIISYQKTVYCALLFKSTRNKRYDAVAKRRELWM
 SLITTKC"

promoter
 complement (128432)
 /note="TATA: TATTAAA before BGLF1, potential promoter for
 3.0kb late RNA"

misc_feature
 128848
 /note="BAM: Bam HI C/D"

promoter
 complement (129054)
 /note="TATA: TATTTC before BDLF4, potential promoter for
 3.7kb early RNA"

CDS
 129188..130348
 /note="BDRF1 reading frame, homologous to RF 42 VZV and
 spliced gene in HSV (Costa et al, 1985). Spliced from
 BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
 1.9kb early RNAs. Possibly also 1.8kb early RNA."

promoter
 129374
 /note="TATA: TATAAGC"

promoter
 complement (129377)
 /note="TATA: TATAAG"

misc_feature
 129413
 /note="DONOR: GTGGTAAGT possible donor"

misc_feature
 130347
 /note="polyA signal: ATTTAA"

misc_feature
 complement (130359)
 /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
 2.3kb late RNA and 3.2kb late RNA"

CDS
 complement (130362..131066)
 /note="BDLF3 late reading frame 9kXNT/S; NCBI gi: 833012"

/codon_start=1
 /translation="MAHAKACAGVAMILICETSLWTSSTSGSSTASAGNVGTGAVT
 TPSPSASPSTNQSTLTITTSAPITTTAILSTNTTTTFTGTTVPVPTTSNSTINV
 TKTVTAQNTATEAGTGTGTVSTNVTSSSTTSATTRITNATTLAPTLSKSGTNSA
 TKTTAEPLTPVDERQPSGLPLMTLVFVGLTFLMLILIFAAGLMSAKNRPLDEAL
 LNAVTRDPSLYKGLV"

promoter
 complement (131104)
 /note="TATA: TATAAAA EEL4 late promoter before BDLF3,
 gives 0.9kb late RNA"

```

CDS
/note="complement (131127..132389)
/note="BDF2 late reading frame; NCBI gi: 833013"
/codon_start=1
/translation="WDQQAIVEHGTIVSHITISREEDGVVHRRVIVASGERVVEFYKAP
APRREGRASTFHDFTVPAAAAVGPEPEPHFPMPIHANGGETKNTDQDQNOT
TRTRTNKAEEAEMDDTMASGCGORAPISADLLSSLTGMAAMAPSMKSEVVC
GERMFKEDVDGEAETLAEPFRCFMLSFVYIYCCYLAFLALLACFNPFLFSPMP
VGAKLRKGDDGVPLSTGYCNPFCVKYTLIPAVVINVTYPNNDTDSHGGEFE
AAALHVAALFESGPNLOARVNNRNTFNTRASGRVERRLVQDMQVILASAVVMHHH
CHVETVYVDCVGPEEGTPTPCFQDVIAERPSLVNTCTAPLKTSVKGENWSGAAGGM
KRKQCKVRDLRDSFPAYLEEVMYVMWQ"
132266
/note="TATA: TATAAAA"
complement (132400..133305)
/note="BDF1 late reading frame, poor homology to RF 41
VZV; NCBI gi: 833014"
/codon_start=1
/translation="MDLKVWVSSRLYTDEIKAMQORTGCCILPLASTHGTQNVQGLG
LQGVYSLETVPDVSNYNYLSDCTLAVIDEVSVDLSILTKIVPGQTVIAKQKYPFFQ
WHGTGSLWMPVFCGEHATVKLESNDVDIVFPMVLPTIAEVLQKILLNWNYSRV
MQAPGNADMIDVAMHIGSVSYLGHYETALPEVPGPLGLALLDLNLSYFCIMWTLPR
ASMRIVRGLIRHEHDLNLFQEMVPEIARIDLDLSDVADLSRNRVNMVTLQSLAS
LFNLGRLTAAYSQETLTATCWLR"
complement (132416)
/note="TATA: TATTAA before BDF2, likely promoter for
2.3kb late RNA"
complement (133312)
/note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
complement (133321..137466)
/note="BDF1 late reading frame, homologous to RF 40 VZV
and major capsid protein of HSV; NCBI gi: 833015"
/codon_start=1
/translation="MASNEGVEVRPPYLTVDADLLSNRQSAEGLFHSFDLLGKD
ARAGKFEVLGGVYNAIQYRLETAIVSCVNTFEKDLSRMDGKIQFRIISVPTI
AHGGRSRKORTFIVKNCCHKHILSTEMELSMLEILHISIPETVEAYVCAVKT
VASAQFGVDALGERLINTVLSVKLRHAPPHEILQTLADPTFERGESKTVKSDLLAM
FKHLLHEFLDRAENMGSGFSQYVRSLSEWAAVSGESVIGKGVSTYTKAGGEPV
GGVIFVNDVIRLQTLTLEEADNQIMGFSSTYASFVVRGENLTVAVSYGRVMTFEHF
MARIVDEKXGSKSDLPAAVAGVEDQRPVISAAYIKLGNHAAVESLQKMYNDTQ
SPFLNRRMQSYTFEPVGLFNPFPYTTSAAIKMDLNPQQPLPVEAWIVNKNLLIIF
NIQAALVKLCHPLRHPARTLNLNAPAPDRDRETSYLSQRRNHNHNVIVIDEYD
NKYAAPTVDIAKGLPTEDFLHPSNYDLLELHPLYDVIYGRDAGERARRHVRHL
MVGNIPTLAPAFQZAGQGFETATSLAHVQDAVETQDPTATYATPAFFYVEA
MIRGFEEKFVMNVPLVSLCINTYERSGRIAFNSFMTKICRHGNNAISKEAYSM
YRKIYGEILIAEQALMAGLSDVGVGQVYVSCALLDPLPVAYTDIFTHLLTVS
DRAPQIILGNEYADTLAOPQ IERVGNDEMAAQFVALYGYRVNGDHDHDFRLHLP
YVDEHADVLEKIFYVFLPTCTNACMGGLGVDFQHVAGTLAYNGPASFSHFTREDGI
LONLENGTIRLELLEISDLRPTVGMIRDLSAFMTCPTFTRAVRVSDNDVTQQLANP
ADKRETEQTVUNGIVAFASERTAVTQCLEHAIPHEMYGDPVAAVTHQDVAFTVM
RNPQQRAREANRPEQLAEYREHRSPMGKIAECLPSLSISGMTAMHKMSPAY
IAQAKLHPGAMTVTRTDEILSENILFSSRASTSMFTCTPNVSRREARVDAVTEV
HHEWASIDTGLSYSTMTPARVAALITDNGIHTQDDEFSFPAEAFGNQVNDYIKAVY
CAQRNCTLLRPRTYIAGTMNVNAGLCHQQRATCEIIVTPVADVAIFQKSNPRG
RAACVNTGHEVYAGELIYDHSRDPDAEYRSVWNPASQLSGDITWYNSSTYQT
AVPGLYSPCRAFFKEELLRNGLNWNVEYSQRLGHPATSNTEQFVVIAGTQVF
LEQPCSFLOEAFALASASRALIDEFMSVKQTHAP IYGHYIIEEVAPVRRI LKFGNK
VF"
complement (133332)
/note="DONOR: AAGGTGGTT possible donor"
complement (133352)
/note="TATA: TATTAA before BDF1"
complement (133386)
/note="TATA: TATATAA"
135178
misc_feature

```

```

promoter
/note="polyA signal: AATAAA"
135394
/note="TATA: TATAAGT"
136624
misc_feature
/note="polyA signal: AATAAA"
136868
misc_feature
/note="BAM: Bam HI D/c"
complement (137710)
/note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
4.5kb late RNA"
137857
/note="TATA: CATAAAC"
137862..139715
/note="BcRF1 reading frame"
138019
/note="BAM: Bam HI c/b"
139352
misc_feature
/note="BAM: Bam HI b/T"
139642..140916
/note="BTRF1 reading frame. Northern blots detect 0.95
late and 3.8kb early RNA"
complement (140902)
/note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
complement (140916..143036)
/note="BXLF2 late reading frame, encodes gp85; homologous
to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV);
NCBI gi: 833016"
/codon_start=1
/translation="WQLICVFCILVMEVGAASLSEVKLHLDIEGHASHYTIPTWELM
AKVPGLSPEALREAVNTEDLASMLNRYKLIYKTSCTGLIALAEPVDIPAVSGSNQV
DASKVHPGVISGLNSPACILSAPLEKQLFYITGTMIPNTPHSHVYVYQLRCHLSYAL
SINGDKQYTGAMTSKFLMGTYKRVTEKDEHVLISLFGKTKDLPDLRGPFSYPSLTS
AOSGDSYLVITVTVHYANFHNYPVNLKDMESRAVMTAAASYAVYLQKLVILLEKMG
GCREFLEDITLTTMFVSVAFFVGHAVGTCGVDLRLAKSFFELTVTKDIIIGI
CGATVKGMSYGLERLAAMLATVKEELGHLTTKEQYALRALATVGYPKAGVYSGL
IGCATSVLSAYNRHP LFQPLHTVWRETFLIGSHVVIRELRNLVTTQCPNLALYQLLS
TALCALEIGEIVLGLAIGSLGFSYLSLRFDLTRKLLSMAPQEAATLDQAASN
AVDGF LGRLSLEREDRAHLPAKYCDRLDKVLMIP LINVTFIISDREVRGSALY
EASTYLLSSLSLSPVMKNCQAGVAGEPRQIPKIQNFTRTOKSCIFCGFALLSYDE
KSGLETTIYITSQEVQNSILSSNFEDNLVHLLLTGTGTVEIACGLVEERAHVWL
AIIYFAFALGIFLVHKIVMFFL"
140970
/note="polyA signal: AATAAA"
complement (141286)
/note="polyA signal: AATAAA"
142589
/note="TATA: CATAAAA"
142740
/note="BAM: Bam HI T/X"
complement (143038..144861)
/note="BXLF1 early reading frame, thymidine kinase
(Littler et al, 1986). Weak homology to RF 36 VZV and HSV
thymidine kinase. 4.0kb early RNA presumably encodes the
TK. Also a 2.2kb late RNA here. NCBI gi: 833018"
/codon_start=1
/translation="WAGFPCKEAGPPGWRKKQCEDESPENERHENFYAETDDFAPSVL
TFTGSDSGAGEDDDGLYQVTHWPPLMAPTGLSGERVPCRTQAATNTNGSPSRH
TSCPTLPRGAQPPAPAHQKPTAPTPKPSRCEGPKTDPDFSWFRKTSCTEGGAUST
SRSFYKQKGFEEGLAGLDDKSDCESEDSNFRPSSHSAIKQKNGCKGPKSGLFEH
HVPVTKACSLFEGAPGVGKTMNLNHLKAVFGDLTIWVPMRYWTHVYENAIKAWH
KNVTSARHGREDTSAEVLCAOMKFTTTPFVRLASRKRSLLVETSGARSVAPLDCHLHLD
RHLLSASVEFPLMLIRSQLLSYDFEIQVLAFTADPGDITVMMKLNVEENRRKKRG
RKHESGLDGYLKSNDAYHAYVCVAMLITQYFAPEDIVKVACGLTITITVCHQSHTPI

```

```
IRSVAEKUYRNSIFSVKEVIQFPRADAVLLEVCLAFTRTRLAYLQFVLVLSFQDD
LPGCWEIYMAQNAKPAIRSOFFDMAGLSKVISDFERNDR"
complement (143310)
/notes="TATA: TATAACA ECL2 late promoter before BXLF2,
143608
gives 2.5kb late RNA"
/notes="polyA signal: AATAAA"
144791
/notes="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
BXRF1"
144860..145606
/notes="BXRF1 late reading frame, homologous to RF 35 VZV.
Basic (core?) protein. NCBI gi: 833017"
/codon_start=1
/translation="MDPTRGICALSTHDLAKEFHSLPARKAKGRHLRCYSKILLSIK
SMEQLASFSLPPGPTTDFRLFFVETLGRRTADVVALQF YPRCYIVEFKTAMST
ANQSVTRKAQRLEGTALQCDANFLRTSCPPVLGSGLEVLALVFNQKRSLETLQV
EFPALGQKTLPTSTGLMLLSRWQDGLARLDRPRTAQGHPRTHVGRKPSQLTA
RVPRSARAGRGRKKQGVAGVQVCPQAK"
144862
/notes="BAM: Bam HI X/V"
144945
/notes="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
complement (145135)
/notes="TATA: TATAACA before BXLF1"
145302
/notes="TATA: TATTTAA before BVRF1, potential promoter for
1.9kb early RNA"
145416..147128
/notes="BVRF1 early reading frame, homologous to RF 34 VZV;
NCBI gi: 833019"
/codon_start=1
/translation="MALSGHVLIDPARLPRDTGCELMWAPSIRNSLVSPEALEAER
EAKRSEKORCAQVKKRNLRLVEDIGMDHDLARAEIIRODLDVAVFSDGLSNQ
VRSPTGCSAPAPSPSPAPQFRLTGNQAVYSISPTDPP LMAVGS LAQTLGCLNY
GNIOWVPSFGPWVRTMSANAMORRVEPKLRGNLNTSVSLKLMTEVAVLEGTQ
DFFSDVRHLIDQALLISVAYLLQGSSSHQRRPLPASREELLEGPESLEKIIDL
RAKSPGNGRILTSNGKEARQSIAPLNROQYFETFDNKNYLFVAGLPLTAAAL
NVP CAGGRDIDLVRITANOIFGEDVPVFSHOMNLRVLGALAEALMIVTLCETANIA
EAATRLHLSLLQAMQRRKPPAMASAGMPGAYPVQTLFRGELFRE IWAHVVRPTVA
ADPQASTISLPGLVLALEIKMDQNP SHYAINLTGQKDTLFEINQKLDJFHPA
AMLAARTQLRLAFEDGVGVALGRPSMLAAREILLERQFSASDDYDRLVFLTUGYIASP
VAP"
complement (146926)
/notes="polyA signal: AATAAA"
147167
/notes="DONOR: AAGGTAAAT possible donor"
147170
/notes="polyA signal: AATAAA, 3' end of 2.4kb late and
1.9kb early RNAs"
147721
/notes="TATA: TATTAT before BVRF2, potential promoter for
2.1kb early RNA"
147927..149744
/notes="BVRF2 early reading frame, N-terminus homologous to
RF 33 VZV; NCBI gi: 833020"
/codon_start=1
/translation="MWQADSVVVGCFVFRPDAPPKQACLHLDPLTVKSQLPKKPIPL
TVEHLDPADVSGVGLYQSRAGLESAASITSCDFLSLIDSIYHDCDLAQSRPLPRE
PKVEALHAWLSLASLHPDIPQTTADGKLSFFDHVSICALGRRGTAVYGTDLA
WLKHFSLEPSTAAQENDANAARESGCPEDHPLPLTKLIKAIADAGFLNRNVEVL
RQDRGVANIPAESVLKASDAPDLQPKALQPPPPASTDPATMLSGNAGEGATACGS
AAAQQLISVPRNFTETLIQTNLDNKPFPQTELPYAAPLPFPFSHOAIATAPSYGPAG
AVAPAGGYTSPGCVYAGAGGDP GAF LAMDAHTYHPHPHPAYFGLGCLFGPPPPV
```

```
PPYSGHRLADYVPAPSRNKRKRDRPEDEEGGLFPGEDATLYRKDIAGLSKSVNEL
QHTTALARETISYGTGVCYQCPQCYTHSGPYGFQHQSYEVPRVYHPPPPTS
HOAAQAPPPCTQAPAEAHCVAEСТИPEAGAACNSCPREDTNPQPTTEGHHRGKLV
QASASGVAQSKEPTTPKAKSVSAHLKSI FCEELLNKRVA"
148007
/notes="BAM: Bam HI V/d"
148620
/notes="TATA: TATTTAA ECR1 late promoter before Bdrf1,
gives 1.2kb late RNA"
148707..149744
/notes="Bdrf1 reading frame; this is the C terminus of
BVRF2; NCBI gi: 833021"
/codon_start=1
/translation="MLSGNAGEATACGSAAGQQLISVPRNTFTLLQTNLDNKRPP
RQTP LTYAAPLPFP SHQAIATAPSYGAGAVAPAGGYTSPGYYAGAGGDPGAF
AMDATYHPHPPPAYFGLGFPVPYYPYSGHRLADYVPAPSRNKRKRDPPE
DEEGGLFPGEDATLYRKDIAGLSKSVNELQHTLAALRETTLSYGTGVCYQCPQC
YTHSGPYGFQHQSYEVPRVYHPPPPTS HQAAQAPPPPTQAPAEAHCVAEСТИPE
AGAACNSCPREDTNPQPTTEGHHRGKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
FCEELLNKRVA"
149115
/notes="BAM: Bam HI d/I"
149727
/notes="polyA signal: AATAAA, 3' end of 2.1kb early and
1.2kb late RNAs"
complement (149758)
/notes="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
late and 1.8kb late RNAs"
complement (149779..150525)
/notes="BILF2 late reading frame 11XNTT/S; NCBI gi:
833022"
/codon_start=1
/translation="WTHLVLLCCCVSGCAFFSDLVKEENVTAHAGARNVLTCSVPS
NESRSRELGRGTGQQLPLAVATYNNGTHITNGGYSLTLEWVNDSTNSLLI
PNVTLAHGYTNTVLNGSVASGVNYSAGEDDQYHANRTLTORMHLTVIPATT
IAPTTLSHTTSTSHRHRPVSQRPTKQVTLGPP IDPWRPKTTVWHALLLITCA
VAPVLLIIIIISCLGMLAGRRRKGWPL"
complement (150571)
/notes="TATA: TATTTAT before BILF2. Potential promoter for
1.0kb late RNA."
151236..151618
/notes="repetitive sequence 3X"
151767
/notes="polyA signal: AATAAA"
complement (151780)
/notes="TATA: CATAAAA"
152012..152013
/notes="DEL: B95-8 deletion with respect to Raji"
complement (152161..153099)
/notes="BILF1 reading frame, membrane protein, 3xNXS/T;
NCBI gi: 833023"
/codon_start=1
/translation="MLSTMAPGSTVGTUVANWTSVNATEDACTKSFAFLSGMTSLLL
VLLIITLAGITLFIIVRKLIHVHMDVWLIALIELLMLVGLMKIQFSTGCLLTQN
MFLGLMCSVWTHLGMALEXTIALFSRTPKRTSHRNVCGLYLMGVFCVILLIILIT
MGDANINRGPNMCREGTTKGMTAVQGLKAGCYLLAAVLIVLTVIIWKLRTRFEG
RPRGLNVTFTGLICAFSWMFLSLPILFLGEAGSIGFCTESLVARYYPGPAACIAL
LLIILYAMFSFHFMDSLKQVTVTARYPRVPQSQT"
152230
/notes="TATA: CATAAAA"
153259
/notes="polyA signal: AATAAA"
153637
/notes="HPN: 22bp 2-fold symmetric"
```

```

misc_feature
/notes="DONOR: AAAAGTCAGC possible donor"
complement(153690)
CDS
/notes="BALF5 DNA polymerase (early), homologous to many
DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA
apparently encodes BALF5, RNA ends unknown. NCBI gi:
833024"
/codon_start=1
/translation="MSGGLFYNPFLRPNKGLLKPKKEYLRLIPKCFQTPGAAGVDV
RGPPPLCFYODSLTVVGGDEGCKGMMWRQAEQGTARPEADTHGSPIDFHVYDILET
VYTHEKAVIPESKQGVVPCGIVIKLGRKADGASVCNVFQQAYFYASAPQGLD
VEFAVLSAKGSTDRPTCRVSEKVTFRSINGHCAGDYHKITLISHPNSECHVAT
WADQKGRIFEAENVDAIRFVINDVFTFGWYSCRRAPRLQHRDSYAELEVDCEVG
DLVSREDSWPSQLAFDIECLGEGEPPATNEADLLIQJSCVLMWSTGEEAGYRR
LLTLGTCDTEGVEYVEPSELDMLYAFQLIRDSLVEITGVNANFDMPTLIDRA
RHIYSINPASKGIRAGGVCVRPHDCKGFLRANTKYRITGLIPIDMYAVCRDKLS
LSDYKLDTVARHLLGAKEDVHYKEIPRLFAAGPGRRLGMVCQDSALWMDLHIF
VHVVEAETAKIAHIPCRVLDGQQIRVFSCLLAAQKENFTILPWPASASRDGYGA
TVIQPLSGFYNSPVLVDVDFASLYPSIIQAHNLCSYTMTPGEEHRLAGLRGDEYSEF
RLTGGYVHFVKVHVESFLASILSWLAKRKAIKKLLAACEDPQRQTILDQQLAKC
TCNAVYFTGVANGLEFPCLSTAETVTLQGTMRERAKAFVIALSPANUALAPSPAW
APLNPEQGLIYIGDTSDFTECRGFSESETLRFADALAAHTRSLSFVAPISLEAKT
FSCMLITKERYVGVLTIDCKTLMKGVELYKRTACKFVQTCRRVRLDVLADARVKEA
SLASHRPFQESFTQGLPVGLPVLPIIDNOAVTDLREGRVPMGELCFSTELSRKLSYK
STQPHLAVYQKGFVERNEELPQIHDIRIOYVTFVFGKGVGARKTEMAEDPAYAERHGV
PVAVDHYDFKLQCAANIILQCLFDNNSGAALSVLQNFATRPFF"
154747
misc_feature
/notes="BAM: Bam H1 I/A"
complement(156707)
misc_feature
/notes="polyA signal: AATAAA; 3' end of 2.5kb late (gb) RNA
and 1.8kb late RNA"
CDS
complement(156749..159322)
/notes="BALF4 late reading frame 9xNXT/S homologous to HSV1
glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31
VZV (gpII); NCBI gi: 833025"
/codon_start=1
/translation="MTRRVLVVVLLAALACRLGAQTPEQPAPPATTVQPTATROQT
SFFPRVCELSHGDLRFSSDIQCPFGTRENHTEGLMFKVKNIIIPYSFKRVSYTKI
VTNILLYNGVADSVTRNEEKEFSYDSYETQMDTIYQCVNAVKMTKGLTRVYVDRI
GYNITVNLKPTGGLANGVRVYASQTELDAPGHLITRYTRTYNCLITDMAKSNP
FDFEFTTLCQTVKESPFYDGKRNKETFHERADSFHRTYNTKYIVDYNRGTNPQGERAF
LDKCTVTLISWKLNRNTAYCPLQHMOTFDSTIATETGKSIFHVTDEGTSFSFVNTTVGI
ELDPAKCEIEQVNMKHEKYEAVDQRYTKQEAITYFTISGGLLAWLPILTPRSLAT
VKNLTELTPYSSPSPSPAPSAAGSTPAAVLRRRRRDAGNATTVPVPTAPGKSL
GLNNPATVQIQFAYDSLAROINRMGLDLARAWCLEQRQNMVRLTKINPTVMSS
IYKGAVAKRILGDVIVSQCVPVQATVTLKSNRVPGSEWICYSRPLVVSFSDINTK
TYEQGLGTNEIFLTKKTEVCQATSQYFQSGNEIHVNDYHHFRTIELDGIATLQT
FISLINTIDFISLEIYSRDRQASNVDFLEGIFREYNFOAQNIAGLRQIDLNAV
SNGBNQFVGLGELMDSLSGVQCSIINLVSIVTGGLESSLVSCFISFFKNPFGCMLIIV
LVAGVWLIVSLTRTRQMSQPVQMLYPGIDELAQHASGEGPGINPISKTELAQIM
F"
158204
promoter
/notes="TATA: TATAAAA"
complement(159312..161678)
CDS
/notes="BALF3 reading frame"
complement(159370)
promoter
/notes="TATA: TATTTAA ECL1 late promoter before BALF4,
gives 2.5kb late RNA"
160966
misc_feature
/notes="polyA signal: AATAAAA"
160990
misc_feature
/notes="polyA signal: ATTTAAA"
misc_feature
complement(161013)

```

```

CDS
/notes="polyA signal: AATAAA, presumed end of 3.9kb early
RNA"
complement(161384..164770)
/notes="BALF2 early reading frame, homologous to RF 29 VZV
and major DNA binding protein HSV. 3.9kb early RNA; NCBI
gi: 833026"
/codon_start=1
/translation="MQGAQTSEDNLGSSQPGPCGYIIFYPLATYPLREVATLGTGYA
GHRCTVWPLLGGITVEGFSINVKALHRRDPDNCGLLEATSYHRDIYVFNHMHVPPPI
FECPGLEALCGTREVTCYDAYSALPRESKPGDFPEGLDPSAYLGAVAITAEKFER
LYSGNLVAIPSLKQEVAVQGSASVRPLYDKVEFPEGVPLQRFYNSDLSKOWHEALY
TGLAQALRRVRKGLFELLESKQLQQAQKAVAPLKEFPASTISHPDSGALMIVDSA
ACELAVSAPAMLEASHETPASLNDVSWPLEADCEGPEARVAALHRYNASLAPHVSTQ
IFATNSVLVYSGVSKSTQGGKESLFNSYTHGLGTQAEQWDPGRPCFSMGGGDV
TCTNGPGRAYAVELHYAASFNNLARYAYIQFQQGKSLTPVETGSYVAGAAAS
PMCSICEGRAPAVCLNTLFRLRDRFPVMSTORRDPYVIGSAGSYNETDFLGNFLN
FIDKEDQQRPDDEPRYTYWQLNQMLLERLSRLGIDAEGKLEKEPHGPRDFVMKFDV
DAADVAEVQFVNSMAKNNTIYKDLVSKSYHVWQYSCNPAQPCIFITQLFVRSLLT
ILQDISLPICMCYENDNPGCGPPEMLKGHYQTLCTNFRSLAIDKGYLTAKCAKVH
GEPTCDLPDLAALQGRVYGRRLPVMSKVMALCPRNIKIKNRVYFTGERAALQNSFI
KSTTRENYYIINGPYNKFLATYHKTLFPDTKLSSLYLHMNFSSRRSVVPVSGSAEY
SDIALVDDGSGRAHEESNVDDVPGLNVITYAKQRLNNAILKACGQTFYISLIQGLVP
RTQSPARDYPHVLGTRAVESAAAYAEATSSLTATTVVCAATDCLSQVCKARVWTLR
VTINQYVGNNGNQIFQAGNLGTFMGRVDNRNLIQAPCAGLRQAGGSMRKKKVFAT
PTLGTITVKNRTQAAATYIENIRAGLEAIIISQKQEDCVFVNVCLNVDAMGEACASLT
RDAEAYILGFSVLADSVLETLATASSGIEWTAEARDFLEGVWCGPQAAQDNFISV
AEPVTSQASAGLLLGCGGSGGRRRRRLATVLPGLV"
163978..166635
misc_feature
/notes="DEL: deletion in Raji"
complement(164814)
promoter
/notes="TATA: CATTTAA before BALF2, presumed promoter for
3.9kb early RNA"
misc_feature
complement(164851)
CDS
/notes="polyA signal: AATAAA"
complement(164855..165517)
/notes="BALF1 early reading frame, 0.7kb early RNA; NCBI
gi: 833028"
/codon_start=1
/translation="MNLAIADSPHPGLASYTILPRPFYHISILKPVSWPDETMRPAKS
TUSVTRTPVEAWVAPSPDDKVAESSYLMFRAMVAVFTRDEKDLPLPALVLCRLKA
SLRKORKIYAEALACTADIGCKDTHVRLIISVLRVAVYNDHYWRLRWILCYTVVFA
VRNYLDDHKSAAFLVGAIAHYLALYRRLWTFARLGGMPRLRRQFPVTTWALASLDTFLK
SL"
165442
promoter
/notes="TATA: GATAAAA"
165466
promoter
/notes="TATA: TATAAGA EDRI early promoter before BALF1,
gives 0.8kb early RNA"
165504..166169
CDS
/notes="BALF1 reading frame; NCBI gi: 833027"
/codon_start=1
/translation="MARFIAQLLLILASVAAAGQAVTAFGLERVTLTSYVRRVSLGPEI
EVSWEKLGPEEQVLIIGMHHDVIEIENPFRRGFDIHRSANTFFLVVTAANIHDQNY
LCRMKLGTEQVKEHLSVVKPLITLSVHSERSQFPDFSVLTVTCTVNAFPHPHQWIM
PEGVEPAPTAANGGVNKEKDSLSVAVDLSILPKPWHLPVTCVGKNDKEEAHGYYVSGY
LSQ"
complement(165713)
promoter
/notes="TATA: TATAAAG before BALF1"
166165
misc_feature
/notes="polyA signal: AATAAA 3' end of 0.8kb early RNA.
Also 1kb late RNA in this region."
166469..166475
promoter
/notes="TATA: TTATTTT"
mRNA
166498..166916

```


sapiens (clone library: Kai Wang's) (tissue library: ATCC 1521)
DNA; Homo sapiens DNA; Homo sapiens DNA; Homo sapiens
sapiens (clone library: Kai Wang's) DNA; Homo sapiens (clone
library: Kai Wang's) DNA; Homo sapiens (clone library: YAC 234
A6F6) DNA; Homo sapiens (clone library: YAC 234 A6F6) DNA; Homo
sapiens (clone library: YAC 234 A6F6) DNA; Homo sapiens DNA; Homo
sapiens DNA; Homo sapiens DNA; Homo sapiens (clone library: YAC 234
A6F6) DNA; Homo sapiens (clone library: YAC 234 A72B3) DNA; Homo
sapiens (clone library: Kai Wang's) DNA; Homo sapiens (clone
library: YAC 199 B157C9) DNA; Homo sapiens (clone library: YAC 234
A72B3) DNA; Homo sapiens (clone library: Kai Wang's) DNA; Homo
sapiens (clone library: Kai Wang's) (tissue library: ATCC 1521)
DNA; Homo sapiens (clone library: Kai Wang's) DNA; Homo sapiens
DNA; and Homo sapiens (clone library: Kai Wang's) DNA.

Homo sapiens

ORGANISM

Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.

1 (bases 1 to 684973)

Rowen, L., Koop, B.F. and Hood, L.
Sequence of the human T cell receptor beta locus
Unpublished (1994)

2 (bases 368065 to 442094)

Slightom, J.L., Siewieniak, D.R., Sieu, L.C., Koop, B.F. and Hood, L.
Nucleotide sequence analysis of 77.7 kb of the human V beta T-cell
receptor gene locus: direct primer-walking using cosmid template
DNAs

JOURNAL	Genomics	20	(2), 149-168	(1994)
MEDLINE	94292194			
COMMENT	NCBI gi:	540583		
FEATURES				
source		Location/Qualifiers		
		1..684973		
		/organism="Homo sapiens"		
		/note="human"		
source		1..37031		
		/note="(vector PME15A)"		
		/organism="Homo sapiens"		
		/cell_line="CGM1"		
		/getmline		
		/sequenced_mol="DNA"		
		/clone="K41A"		
		/clone_lib="YAC D49H4"		
		/citation=[1]		
repeat_unit		complement(1..241)		
		/note="putative"		
repeat_unit		/rpt family="Alu"		
		2429..2988		
		/note="putative"		
repeat_unit		/rpt family="LINE 1"		
		7997..8275		
		/note="putative"		
repeat_unit		/rpt family="Alu"		
		8317..8427		
		/note="putative"		
repeat_unit		/rpt family="MER33"		
		10326..10600		
		/note="putative"		
repeat_unit		/rpt family="Alu"		
		complement(23629..23893)		
		/note="putative"		
repeat_unit		/rpt family="Alu"		
		complement(25987..26233)		
		/note="putative"		
repeat_unit		/rpt family="Alu"		
		28056..28507		

repeat_unit	/note="putative" /rpt family="LINE 1" (35482..353482)..353798 /note="putative" /rpt family="MSTA" 36258..72769 /note="(vector PWE15A)" /organism="Homo sapiens" /cell line="CCM1" /germline /sequenced_mol="DNA" /clone="K35" /clone lib="YAC D49H4" /citation=[1] /note="putative" /rpt family="Alu" complement(41973..42235) /note="putative" complement(42818..43079) /note="putative" /rpt family="Alu" 45446..45721 /note="putative" /rpt family="Alu" complement(49022..49282) /note="putative" /rpt family="Alu" complement(51129..51357) /note="putative" /rpt family="MLT1A" complement(52099..54780) /note="putative" /rpt family="LINE 1" 54870..54937 /note="putative" /rpt family="MER25" complement(56562..56788) /note="putative" /rpt family="Alu" 58592..58862 /note="putative" /rpt family="Alu" 60007..60287 /note="putative" /rpt family="Alu" 67296..71811 /note="putative" /rpt family="LINE 1" 70574..70940 /note="putative" /rpt family="THE transposon like element" 71776..107249 /note="(vector PWE15A)" /organism="Homo sapiens" /cell line="CCM1" /germline /sequenced_mol="DNA" /clone="K26" /clone lib="YAC D49H4" /citation=[1] complement(74111..74372) /note="putative" /rpt family="Alu" 77536..77883 /note="putative"
source	
repeat_unit	

```

repeat_unit  /rpt_family="MLTIA"
              complement(78720..78991)
              /note="putative"
repeat_unit  /rpt_family="Alu"
              complement(79236..79567)
              /note="putative"
mRNA         /rpt_family="MSTA"
              complement(join(<79799..79948,80489..80625,81147..81400,
82186..82348,84121..84159))
              /gene="TRYF"
exon         complement(<79799..79948)
              /gene="TRYF"
              /note="putative"
              /pseudo
              /number=5
CDS          complement(join(79799..79948,80489..80625,81147..81400,
82186..82348,84121..84159))
              /gene="TRYF"
              /note="putative"
              /pseudo
              /codon_start=1
              complement(79949..80488)
              /gene="TRYF"
              /note="putative"
              /number=4
exon         complement(80489..80625)
              /gene="TRYF"
              /note="putative"
              /pseudo
              /number=4
              complement(80626..81146)
              /gene="TRYF"
              /note="putative; does not fit consensus"
              /number=3
exon         complement(81147..81400)
              /gene="TRYF"
              /note="putative"
              /pseudo
              /number=3
              complement(81401..82185)
              /gene="TRYF"
              /note="putative"
              /number=2
exon         complement(82186..82348)
              /gene="TRYF"
              /note="putative"
              /pseudo
              /number=2
              complement(82349..84120)
              /gene="TRYF"
              /note="putative"
              /number=1
              complement(82873..83151)
              /note="putative"
              /rpt_family="Alu"
exon         complement(84121..>84159)
              /gene="TRYF"
              /note="putative"
              /pseudo
              /number=1
              complement(85009..85072)
              /note="putative"
              /rpt_family="MIR"
repeat_unit  85538..85773

```

```

repeat_unit  /note="putative"
              /rpt_family="Alu"
              complement(86349..86618)
              /note="putative"
repeat_unit  /rpt_family="Alu"
              complement(90001..90268)
              /note="putative"
              /rpt_family="Alu"
              join(91557..91597,91715..92006)
              /partial
              /gene="TCRBV27S1"
              /note="putative"
              /pseudo
              /codon_start=1
              <91557..91597
              /gene="TCRBV27S1"
              /note="putative"
              /pseudo
              /number=1
              91598..91714
              /gene="TCRBV27S1"
              /note="putative"
              /number=1
              91715..92006
              /gene="TCRBV27S1"
              /note="putative"
              /pseudo
              /number=2
              92007..92013
              /gene="TCRBV27S1"
              /note="RSS heptamer - awaiting approval of new feature
              key; putative"
              92007..92045
              /gene="TCRBV27S1"
              /note="RSS - awaiting approval of new feature key;
              heptamer is at 5' end; putative"
              92014..92036
              /gene="TCRBV27S1"
              /note="RSS spacer - awaiting approval of new feature key;
              putative"
              92037..92045
              /gene="TCRBV27S1"
              /note="RSS nonamer - awaiting approval of new feature key;
              putative"
              complement(92611..92867)
              /note="putative"
              /rpt_family="Alu"
              93469..93478
              /gene="TCRBV22S1"
              /note="conserved decamer"
              join(93552..93600,93689..93986)
              /gene="TCRBV22S1"
              /note="putative"
              <93552..93600
              /gene="TCRBV22S1"
              /note="putative"
              /number=1
              93601..93688
              /gene="TCRBV22S1"
              /note="putative"
              /number=1
              93689..93986
              /gene="TCRBV22S1"
              /note="putative"

```

```

source
/number=2
93930..130694
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="K56"
/clone_lib="YAC D49H4"
/citation=[1]
93987..94025
/misc_signal
/genes="TCRBV22S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
93987..93993
/misc_signal
/notes="RSS_heptamer - awaiting approval of new feature
key"
93994..94016
/misc_signal
/notes="RSS_spacer - awaiting approval of new feature key"
94017..94025
/misc_signal
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(95946..96207)
/repeat_unit
/notes="putative"
/rpt_family="Alu"
complement(98840..99104)
/repeat_unit
/notes="putative"
/rpt_family="Alu"
99921..99960
/repeat_region
/notes="polymorphic"
/rpt_family="microsatellite, gt"
/rpt_unit=99921..99922
/evidence=experimental
101036..101045
/misc_feature
/genes="TCRBV9S1"
/notes="conserved decamer; putative"
join(101150..101198,101315..101609)
V_segment
/genes="TCRBV9S1"
/notes="putative"
<101150..101198
/exon
/genes="TCRBV9S1"
/notes="putative"
/number=1
101199..101314
/intron
/genes="TCRBV9S1"
/notes="putative"
/number=1
101315..101609
/exon
/genes="TCRBV9S1"
/notes="putative"
/number=2
/codon_start=3
101610..101616
/misc_signal
/notes="RSS_heptamer - awaiting approval of new feature
key"
101610..101648
/misc_signal
/genes="TCRBV9S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
101617..101639
/misc_signal
/notes="RSS_spacer - awaiting approval of new feature key"
101640..101648
/misc_signal
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(104481..104664)
/repeat_unit
/notes="putative"

```

```

misc_feature
/rpt_family="Alu"
105669..105678
/genes="TCRBV7S1"
/notes="conserved decamer; putative"
join(105774..105822,105933..106227)
V_segment
/genes="TCRBV7S1"
/notes="putative"
<105774..105822
/exon
/genes="TCRBV7S1"
/notes="putative"
/number=1
105823..105932
/intron
/genes="TCRBV7S1"
/notes="putative"
/number=1
105933..106227
/exon
/genes="TCRBV7S1"
/notes="putative"
/number=2
106228..106266
/misc_signal
/genes="TCRBV7S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
106228..106234
/misc_signal
/notes="RSS_heptamer - awaiting approval of new feature
key"
106235..106257
/misc_signal
/notes="RSS_spacer - awaiting approval of new feature key"
106258..106266
/misc_signal
/notes="RSS_nonamer - awaiting approval of new feature key"
106518..106791
/repeat_unit
/notes="putative"
/rpt_family="Alu"
complement(107940..108191)
/repeat_unit
/notes="putative"
/rpt_family="LINE 1"
complement(110274..110535)
/repeat_unit
/notes="putative"
/rpt_family="Alu"
113509..113518
/misc_feature
/genes="TCRBV5S1"
/notes="conserved decamer; putative"
join(113622..113670,113798..114091)
V_segment
/genes="TCRBV5S1"
/notes="putative"
<113622..113670
/exon
/genes="TCRBV5S1"
/notes="putative"
/number=1
113671..113797
/intron
/genes="TCRBV5S1"
/notes="putative"
/number=1
113798..114091
/exon
/genes="TCRBV5S1"
/notes="putative"
/number=2
114092..114130
/misc_signal
/genes="TCRBV5S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
114092..114098
/misc_signal
/notes="RSS_heptamer - awaiting approval of new feature
key"

```



```

misc_signal 138532..138538
/notes="RSS heptamer - awaiting approval of new feature
key"
misc_signal 138532..138570
/genes="TCRBV7S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
138539..138561
/notes="RSS spacer - awaiting approval of new feature key"
138562..138570
/notes="RSS nonamer - awaiting approval of new feature key"
replace(138563,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(138639,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139298,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139322,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139650,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(139980,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140057..140058,"at")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140126..140129,"ta")
/notes="found in cosmid clone HVB15 in L36190"
replace(140236,"t")
/notes="found in cosmid clone HVB15 in L36190"
replace(140271,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140374,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(140399,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
(140495..140709)..(162290..162390)
/notes="hypothetical location of insertion/deletion
polymorphism, breakpoints correspond to position 4141-4223
in cosmid HVB15 in accession number L36190; putative"
complement(141314..141574)
/notes="putative"
/notes="putative"
/rpt_family="Alu"
141809..141818
/genes="TCRBV13S2"
/notes="conserved decamer; putative"
join(141898..141946,142036..142330)
/genes="TCRBV13S2"
/notes="putative"
<141898..141946
/genes="TCRBV13S2"
/notes="putative"
/number=1
141947..142035
/genes="TCRBV13S2"
/notes="putative"
/number=1
142036..142330
/genes="TCRBV13S2"
/notes="putative"
/number=2
142331..142337
/notes="RSS heptamer - awaiting approval of new feature
key"
142331..142369

```

```

/genes="TCRBV13S2"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
142338..142360
/notes="RSS spacer - awaiting approval of new feature key"
142361..142369
/notes="RSS nonamer - awaiting approval of new feature key"
145518..145557
/notes="polymorphic"
/rpt_family="microsatellite"
/rpt_unit=145518..145519
/evidence=experimental
150631..150733
/notes="putative"
/rpt_family="Alu"
151765..151774
/genes="TCRBV9S2"
/notes="conserved decamer; putative"
<151879..151927
/genes="TCRBV9S2"
/notes="putative"
/pseudo
/number=1
join(151879..151927,152044..152338)
/genes="TCRBV9S2"
/notes="putative"
/pseudo
151928..152043
/genes="TCRBV9S2"
/notes="putative"
/number=1
151952..186019
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="G54"
/clone_lib="Kai Wang's"
/tissue_lib="ATCC 1521"
/citation=[1]
152044..152338
/genes="TCRBV9S2"
/notes="putative"
/pseudo
/number=2
152339..152345
/notes="RSS heptamer - awaiting approval of new feature
key"
152339..152377
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
152346..152368
/notes="RSS spacer - awaiting approval of new feature key"
152369..152377
/notes="RSS nonamer - awaiting approval of new feature key"
155233..155242
/genes="TCRBV7S2"
/notes="conserved decamer; putative"
join(155338..155386,155497..155791)
/genes="TCRBV7S2"
/notes="putative"
<155338..155386
/genes="TCRBV7S2"
/notes="putative"

```

```

/number=1
155387..155496
/genes="TCRBV752"
/notes="putative"
/number=1
155497..155791
/genes="TCRBV752"
/notes="putative"
/number=2
155792..155830
/genes="TCRBV752"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
155792..155798
/notes="RSS_heptamer - awaiting approval of new feature
key"
155799..155821
/notes="RSS_spacer - awaiting approval of new feature key"
155822..155830
/notes="RSS_nonamer - awaiting approval of new feature key"
157060..157320
/notes="putative"
/rpt_family="Alu"
replace(162571..162572,"cc")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162574,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162601,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162622,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162763,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(162969,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
complement(162995..163257)
/notes="putative"
/rpt_family="Alu"
replace(163006,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163460,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163479,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
163490..163499
/genes="TCRBV13S9"
/notes="conserved decamer; putative"
replace(163552,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163562,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
join(163579..163627,163717..164011)
/genes="TCRBV13S9"
/notes="putative"
<163579..163627
/genes="TCRBV13S9"
/notes="putative"
/number=1
replace(163604..163607,"cagc")
/genes="TCRBV13S9"
/notes="putative"
163628..163716
/genes="TCRBV13S9"
/notes="putative"

```

```

variation
/number=1
replace(163641,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163652..163653,"ca")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163669..163671,"ag")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163677,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163680..163681,"gt")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163685,"g")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163690,"c")
/notes="found in cosmid clone HVB15 in L36190; putative"
replace(163705,"t")
/notes="found in cosmid clone HVB15 in L36190; putative"
163717..164011
/genes="TCRBV13S9"
/notes="putative"
/number=2
164012..164018
/notes="RSS_heptamer - awaiting approval of new feature
key"
164012..164050
/genes="TCRBV13S9"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
164019..164041
/notes="RSS_spacer - awaiting approval of new feature key"
164042..164050
/notes="RSS_nonamer - awaiting approval of new feature key"
replace(164142,"a")
/notes="found in cosmid clone HVB15 in L36190; putative"
164352..203210
/notes="(vector PTL5)"
/organism="Homo sapiens"
/cell_line="HeLa"
/germline
/sequenced_mol="DNA"
/clone="H137"
replace(164472,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(164630,"c")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(164983,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165007,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165593,"c")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165698,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165810,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
replace(165949,"a")
/notes="found in cosmid H137 and in cosmid clone HVB15 in

```

```

variation      L36190; putative"
                replace(165972,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(165986,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166193..166195,"ta")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166294,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166310,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166389,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166417..166421,"ccaat")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166428,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166433,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166501,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166508,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166529,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166896,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
variation      L36190; putative"
                replace(166913,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                167102..167111
                /gene="TCRBV6S7"
                /note="conserved decamer; putative"
variation      replace(167136,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
V_segment      Join(167212..167260,167411..167708)
                /gene="TCRBV6S7"
                /note="putative"
                <167212..167260
                /gene="TCRBV6S7"
                /note="putative"
                /number=1
allele          replace(167260,"t")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                167261..167410
                /gene="TCRBV6S7"
                /note="putative"
                /number=1
intron

```

```

variation      replace(167289..167306,"tt")
                /note="found in cosmid clone HVB15 in L36190; putative"
                167291..167336
                /note="polymorphic"
repeat_region  /rpt_family="microsatellite"
                /rpt_unit=167291..167292
                /evidence=experimental
exon            167411..167708
                /gene="TCRBV6S7"
                /note="putative"
                /number=2
allele          replace(167532,"g")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(167633,"a")
                /gene="TCRBV6S7"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                167709..167747
                /gene="TCRBV6S7"
                /note="RSS - awaiting approval of new feature key;
                heptamer is at 5' end; putative"
                167709..167715
                /note="RSS heptamer - awaiting approval of new feature
                key"
                167716..167738
                /note="RSS spacer - awaiting approval of new feature key"
                167739..167747
                /note="RSS nonamer - awaiting approval of new feature key"
                replace(167778,"g")
                /note="found in cosmid H137; putative"
variation      replace(168073,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168168,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168388,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168473,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(168864,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169009,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169470,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169499,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169521,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169567..169569,"caaa")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
                L36190; putative"
                replace(169688,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in

```

```

variation      L36190; putative"
                replace(169885,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170006,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170579,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170646,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(170683,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171468,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171496..171497,"ctt")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171590,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
repeat_unit    complement(171597..171873)
                /note="putative"
                /rpt_family="Alu"
variation      replace(171787,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(171997,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172230,"t")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172297,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172462,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172745,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172762,"c")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(172878,"a")
                /note="found in cosmid clone HVB15 in L36190; putative"
variation      replace(172883..172891,"ccaaagctc")
                /note="found in cosmid clone HVB15 in L36190; putative"
repeat_unit    complement(172985..173195)
                /note="polymorphic; putative"
                /rpt_family="Alu"
variation      replace(173216,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
exon           173318..173614
                /gene="TCRBV30S1"
                /note="putative"
                /pseudo
                /number=2

```

```

allele         replace(173390,"t")
                /gene="TCRBV30S1"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
allele         replace(173532,"a")
                /gene="TCRBV30S1"
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
misc_signal    173615..173621
                /gene="TCRBV30S1"
                /note="RSS heptamer - awaiting approval of new feature
key; heptamer is at 5' end. There is a deletion in the
spacer and 3' end is unknown.; putative"
variation      replace(173623,"g")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
variation      replace(174615..174616,"gct")
                /note="found in cosmid H137; putative"
                complement(174943..177008)
                /note="putative"
                /rpt_family="LINE 1"
                complement(175783..176050)
                /note="putative"
                /rpt_family="Alu"
                replace(176058,"t")
                /note="found in cosmid H137; putative"
                replace(176452,"g")
                /note="found in cosmid H137; putative"
                replace(177403,"t")
                /note="found in cosmid H137; putative"
                replace(178325..178326,"gct")
                /note="found in cosmid H137; putative"
                replace(178623..178626,"gtc")
                /note="found in cosmid H137; putative"
                replace(179131,"a")
                /note="found in cosmid H137 and in cosmid clone HVB15 in
L36190; putative"
                replace(179150,"c")
                /note="found in cosmid H137; putative"
                replace(179294..179295,"atct")
                /note="found in cosmid H137; putative"
                replace(179414,"g")
                /note="found in cosmid H137; putative"
                /note="found in cosmid H137; putative"
                179576..179776
                /note="putative"
                /rpt_family="MER43"
                complement(180768..181166)
                /note="putative"
                /rpt_family="MSTA"
                replace(181040..181043,"ta")
                /note="found in cosmid H137; putative"
                181515..183318
                /note="putative"
                /rpt_family="THE transposon like element"
                replace(181551,"t")
                /note="found in cosmid H137; putative"
                /note="found in cosmid H137; putative"
                /note="found in cosmid H137; putative"
                complement(181892..182155)
                /note="putative"
                /rpt_family="Alu"
                replace(181912,"a")
                /note="found in cosmid H137; putative"
                replace(182049,"t")

```

```

variation /note="found in cosmid H137; putative"
           replace(182640,"a")
variation /notes="found in cosmid H137; putative"
           replace(183279,"g")
repeat_unit /notes="found in cosmid H137; putative"
            complement(183519..183895)
            /note="putative"
            /rpt_family="MSTA"
            replace(183761,"t")
variation /note="found in cosmid H137; putative"
           replace(185410,"g")
variation /notes="found in cosmid H137; putative"
           replace(185928,"t")
misc_feature /notes="found in cosmid H137; putative"
            187054..187063
            /gene="TCRBV31S1"
            /note="conserved decamer; putative"
            join(187166..187214,187324..187616)
            /partial
            /gene="TCRBV31S1"
            /note="putative"
            /pseudo
            /codon_start=1
            <187166..187214
            /gene="TCRBV31S1"
            /note="putative"
            /pseudo
            /number=1
            187215..187323
            /gene="TCRBV31S1"
            /note="putative"
            /number=1
            187324..187616
            /gene="TCRBV31S1"
            /note="putative"
            /pseudo
            /number=2
            187617..187623
            /gene="TCRBV31S1"
            /note="RSS heptamer - awaiting approval of new feature
            key; heptamer is at 5' end. Span of RSS not known.;"
            putative"
repeat_unit 188638..188908
            /notes="putative"
            /rpt_family="Alu"
repeat_unit 189459..189726
            /notes="putative"
            /rpt_family="LINE 1"
repeat_region 189754..189806
            /note="polymorphic"
            /rpt_family="microsatellite"
            /rpt_unit=189754..189755
            /evidence=experimental
repeat_unit 190067..190346
            /notes="putative"
            /rpt_family="Alu"
source 193517..230319
        /note="(vector PTL5)"
        /organism="Homo sapiens"
        /cell_type="sperm"
        /germline
        /sequenced_mol="DNA"
        /clone="H18"
misc_feature 195433..195442

```

```

V_segment /gene="TCRBV13S5"
           /note="conserved decamer; putative"
           join(195522..195570,195662..195956)
           /gene="TCRBV13S5"
           /note="putative"
           <195522..195570
           /gene="TCRBV13S5"
           /note="putative"
           /number=1
           195571..195661
           /gene="TCRBV13S5"
           /note="putative"
           /number=1
           195662..195956
           /gene="TCRBV13S5"
           /number=2
           195957..195995
           /gene="TCRBV13S5"
           /note="RSS - awaiting approval of new feature key;
           heptamer is at 5' end; putative"
           195957..195963
           /note="RSS heptamer - awaiting approval of new feature
           key"
           195964..195986
           /note="RSS spacer - awaiting approval of new feature key"
           195987..195995
           /note="RSS nonamer - awaiting approval of new feature key"
           replace(198022,"t")
           /note="from cosmid H18; putative"
           198980..198989
           /gene="TCRBV6S1"
           /note="conserved decamer; putative"
           join(199093..199141,199252..199549)
           /gene="TCRBV6S1"
           /note="putative"
           <199093..199141
           /gene="TCRBV6S1"
           /number=1
           199142..199251
           /gene="TCRBV6S1"
           /note="putative"
           /number=1
           199173..199202
           /note="polymorphic"
           /rpt_family="microsatellite"
           /rpt_unit=199173..199174
           /evidence=experimental
           199252..199549
           /gene="TCRBV6S1"
           /note="putative"
           /number=2
           199550..199588
           /gene="TCRBV6S1"
           /note="RSS - awaiting approval of new feature key;
           heptamer is at 5' end; putative"
           199550..199556
           /note="RSS heptamer - awaiting approval of new feature
           key"
           199557..199579
           /note="RSS spacer - awaiting approval of new feature key"
           199580..199588
           /note="RSS nonamer - awaiting approval of new feature key"
           join(200895..200939,201077..201349)
           /partial
           V_segment

```

```

/genes="TCRBV32S1"
/notes="putative"
/pseudo
/codon_start=1
<200895..200939
/genes="TCRBV32S1"
/notes="putative"
/pseudo
/number=1
200940..201076
/genes="TCRBV32S1"
/notes="5' end does not fit consensus gt; putative; does
not fit consensus"
201077..201349
/genes="TCRBV32S1"
/notes="putative"
/pseudo
/number=2
201350..201394
/genes="TCRBV32S1"
/notes="RSS - awaiting approval of new feature key; spacer
sequence too long; putative"
201350..201356
/genes="TCRBV32S1"
/notes="TCRBV32S1"
/notes="RSS heptamer - awaiting approval of new feature
key; putative"
201357..201385
/genes="TCRBV32S1"
/notes="RSS_spacer - awaiting approval of new feature key;
putative"
201386..201394
/genes="TCRBV32S1"
/notes="RSS_nonamer - awaiting approval of new feature key;
putative"
201395..203959
/genes="TCRBV55"
/notes="putative"
/pseudo
/codon_start=1
<203911..203959
/genes="TCRBV55"
/notes="putative"
/pseudo
/number=1
203960..204083
/genes="TCRBV55"
/notes="5' end no gt consensus splice; putative; does not
fit consensus"
/number=1
204084..204377
/genes="TCRBV55"
/notes="putative"
/pseudo
/number=2
204378..204416
/genes="TCRBV55"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
204378..204384
/notes="RSS_heptamer - awaiting approval of new feature
key"
204385..204407
/notes="RSS_spacer - awaiting approval of new feature key"

```

```

misc_signal 204408..204416
/notes="RSS nonamer - awaiting approval of new feature key"
V_segment
join(206647..206695,206828..207121)
/genes="TCRBV1S1"
/notes="putative"
<206647..206695
/genes="TCRBV1S1"
/notes="putative"
/number=1
206696..206827
/genes="TCRBV1S1"
/notes="putative"
/number=1
206828..207121
/genes="TCRBV1S1"
/notes="putative"
/number=2
207122..207160
/genes="TCRBV1S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
207122..207160
/genes="TCRBV1S1"
/notes="RSS heptamer - awaiting approval of new feature
key"
207122..207128
/notes="RSS_heptamer - awaiting approval of new feature
key"
207129..207151
/notes="RSS_spacer - awaiting approval of new feature key"
207129..207151
/notes="RSS_spacer - awaiting approval of new feature key"
207152..207160
/notes="RSS nonamer - awaiting approval of new feature key"
207152..207160
/notes="RSS nonamer - awaiting approval of new feature key"
complement(207825..208060)
/notes="putative"
/rpt_family="Alu"
209626..209890
/notes="putative"
/rpt_family="Alu"
complement(211625..211815)
/notes="putative"
/rpt_family="Alu"
214553..214562
/genes="TCRBV12S4"
/notes="conserved decamer; putative"
join(214638..214686,214793..215087)
/genes="TCRBV12S4"
/notes="putative"
<214638..214686
/genes="TCRBV12S4"
/notes="putative"
/number=1
214687..214792
/genes="TCRBV12S4"
/notes="putative"
/number=1
214793..215087
/genes="TCRBV12S4"

```

```

/misc_signal
/number=2
/215088..215126
/genes="TCRBV12S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
215088..215094
/notes="RSS_heptamer - awaiting approval of new feature
key"
215095..215117
/notes="RSS_spacer - awaiting approval of new feature key"
215118..215126
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(221378..221646)
/notes="putative"
/rpt_family="Alu"
222287..222296
/genes="TCRBV21S1"
/notes="conserved decamer; putative"
join(222397..222445,222547..222844)
/genes="TCRBV21S1"
/notes="putative"
<222397..222445
/genes="TCRBV21S1"
/notes="putative"
/number=1
222446..222546
/genes="TCRBV21S1"
/notes="putative"
/number=1
222547..222844
/genes="TCRBV21S1"
/notes="putative"
/number=2
222845..222883
/genes="TCRBV21S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
222845..222851
/notes="RSS_heptamer - awaiting approval of new feature
key"
222852..222874
/notes="RSS_spacer - awaiting approval of new feature key"
222875..222883
/notes="RSS_nonamer - awaiting approval of new feature key"
225419..226732
/notes="putative"
/rpt_family="LINE 1"
complement(228277..228621)
/notes="putative"
/rpt_family="MLT1D"
229815..229824
/genes="TCRBV8S4"
/notes="conserved decamer; putative"
join(229930..229975,230075..230372)
/genes="TCRBV8S4"
/notes="putative"
/pseudo
/codon_start=1
<229930..229975
/genes="TCRBV8S4"
/notes="putative"
/pseudo
/number=1

```

intron	229976..230074 /gene="TCRBV8S4" /note="putative" /number=1 230075..230372 /gene="TCRBV8S4" /note="putative" /pseudo /number=2 230204..230630 /note="The gap between cosmid clones H18 and V13G15 was closed by the sequencing of a PCR reaction product." /organism="Homo sapiens" /germline /sequenced_mol="DNA" /clone="H18/G15gap" /citation=[1] 230373..230411 /gene="TCRBV8S4" /note="RSS - awaiting approval of new feature key; heptamer is at 5' end; putative" 230373..230379 /note="RSS_heptamer - awaiting approval of new feature key" 230380..230402 /note="RSS_spacer - awaiting approval of new feature key" 230403..230411 /note="RSS_nonamer - awaiting approval of new feature key" 230482..230750 /note="putative" /rpt_family="Alu" 230528..267156 /note="(vector PWE15A)" /organism="Homo sapiens" /cell_line="CGM1" /germline /sequenced_mol="DNA" /clone="G15" /clone_lib="Kai Wang's" /citation=[1] 233955..234049 /note="putative" /rpt_family="Alu" complement(235992..236357) /note="putative" /rpt_family="Alu" 239619..239628 /gene="TCRBV12S3" /note="conserved decamer; putative" join(239704..239752,239859..240153) /gene="TCRBV12S3" /note="putative" 239704..239752 /gene="TCRBV12S3" /note="putative" /number=1 239753..239858 /gene="TCRBV12S3" /note="putative" /number=1 239859..240153 /gene="TCRBV12S3" /note="putative" /number=2
--------	--

```

misc_signal 240154..240192
/gene="TCRBV12S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
240154..240160
/notes="RSS_heptamer - awaiting approval of new feature
key"
240161..240183
/notes="RSS_spacer - awaiting approval of new feature key"
240184..240192
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(240700..240964)
/notes="putative"
/rpt_family="Alu"
241323..241603
/notes="putative"
/rpt_family="Alu"
243147..243302
/notes="243302; putative"
/rpt_family="Alu"
complement(244346..244691)
/notes="putative"
/rpt_family="THE transposon like element"
complement(246039..246420)
/notes="putative"
/rpt_family="MSTA"
248553..248562
/gene="TCRBV21S3"
/notes="conserved decamer; putative"
join(248664..248712,248805..249102)
/gene="TCRBV21S3"
/notes="putative"
/codon_start=1
<248664..248712
/gene="TCRBV21S3"
/notes="putative"
/number=1
Join(248664..248712,248805..249102)
/partial
/gene="TCRBV21S3"
/notes="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624708"
/codon_start=1
/translation="MCTRLCWAALCILGAEITEAGVAQSPRYKIIEKRSVAFWQNP
ISGHATLYWYQOILGQGPCKLLIQFQNGVDDSLPKDRFSAEIRKGVDSITLKTPAK
LEDSAVYLCASSL"
248713..248804
/gene="TCRBV21S3"
/notes="putative"
/number=1
248805..249102
/gene="TCRBV21S3"
/notes="putative"
/number=2
249103..249141
/gene="TCRBV21S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
249103..249109
/notes="RSS_heptamer - awaiting approval of new feature
key"
249110..249132

```

```

misc_signal 249133..249141
/notes="RSS_spacer - awaiting approval of new feature key"
complement(249302..249573)
/notes="putative"
/rpt_family="Alu"
complement(250307..250649)
/notes="putative"
/rpt_family="THE transposon like element"
252152..252232
/notes="putative"
/rpt_family="MER4"
complement(253936..254362)
/notes="putative"
/rpt_family="MLTID"
255476..255485
/notes="conserved decamer; putative"
join(255591..255636,255736..256033)
/partial
/gene="TCRBV8S5"
/notes="putative"
/pseudo
/codon_start=1
<255591..255636
/gene="TCRBV8S5"
/notes="putative"
/pseudo
/number=1
255637..255735
/gene="TCRBV8S5"
/notes="putative"
/number=1
255736..256033
/gene="TCRBV8S5"
/notes="putative"
/pseudo
/number=2
256034..256072
/gene="TCRBV8S5"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
256034..256040
/notes="RSS_heptamer - awaiting approval of new feature
key"
256041..256063
/notes="RSS_spacer - awaiting approval of new feature key"
256064..256072
/notes="RSS_nonamer - awaiting approval of new feature key"
256774..256996
/notes="putative"
/rpt_family="Alu"
257519..249573
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="X1A"
/clone_lib="Kai Hang's"
/citation={}
261812..262088
/notes="putative"
/rpt_family="Alu"
263528..263921

```

```

/misc_feature
/rpt family="LINE 1"
265632..265641
/genes="TCRBV13S1"
/notes="conserved decamer; putative"
V_segment
join(265721..265769,265862..266156)
/genes="TCRBV13S1"
/notes="putative"
exon
<265721..265769
/genes="TCRBV13S1"
/notes="putative"
/number=1
intron
265770..265861
/genes="TCRBV13S1"
/notes="putative"
/number=1
exon
265862..266156
/genes="TCRBV13S1"
/notes="putative"
/number=2
/misc_signal
266157..266195
/genes="TCRBV13S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
266157..266163
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
266164..266186
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
266187..266195
/notes="RSS_nonamer - awaiting approval of new feature key"
/misc_feature
269768..269777
/genes="TCRBV6S11"
/notes="conserved decamer; putative"
V_segment
join(269879..269927,270043..270340)
/genes="TCRBV6S11"
/notes="putative"
/codon start=1
exon
<269879..269927
/genes="TCRBV6S11"
/notes="putative"
/number=1
CDS
join(269879..269927,270043..270340)
/partial
/genes="TCRBV6S11"
/notes="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624709"
/codon start=1
/translation="MGRILLCWVVLGFLGTDHTGAGVSQSPRYKVKRGRDVALRCDS
ISGHVTLVWYRQTIGQGSSEVLTYSQSDAQROKSGRPSGRFSNERPERSVSTIKIQTE
QGDSAVYICASSLA"
intron
269928..270042
/genes="TCRBV6S11"
/notes="putative"
/number=1
exon
270043..270340
/genes="TCRBV6S11"
/notes="putative"
/number=2
/misc_signal
270341..270379
/genes="TCRBV6S11"
/notes="RSS - awaiting approval of new feature key;

```

```

/misc_signal
heptamer is at 5' end; putative"
270341..270347
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
270348..270370
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
270371..270379
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit
276368..276698
/notes="putative"
/notes="putative"
V_segment
/rpt family="MSTA"
join(277826..277874,277997..278290)
/genes="TCRBV5S6"
/notes="putative"
exon
<277826..277874
/genes="TCRBV5S6"
/notes="putative"
/number=1
intron
277875..277996
/genes="TCRBV5S6"
/notes="putative"
/number=1
exon
277997..278290
/genes="TCRBV5S6"
/number=2
/misc_signal
278291..278329
/genes="TCRBV5S6"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
278291..278297
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
278298..278320
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
278321..278329
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit
281194..281439
/notes="putative"
/rpt family="Alu"
284207..284216
/genes="TCRBV13S6"
/notes="conserved decamer; putative"
V_segment
join(284296..284344,284434..284728)
/genes="TCRBV13S6"
/notes="putative"
exon
<284296..284344
/genes="TCRBV13S6"
/notes="putative"
/number=1
intron
284345..284433
/genes="TCRBV13S6"
/notes="putative"
exon
284434..284728
/genes="TCRBV13S6"
/notes="putative"
/number=2
/misc_signal
284729..284767
/genes="TCRBV13S6"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
284729..284735
/notes="RSS_heptamer - awaiting approval of new feature
key"

```

```

misc_signal 284736..284758
/notes="RSS spacer - awaiting approval of new feature key"
misc_signal 284759..284767
/notes="RSS nonamer - awaiting approval of new feature key"
repeat_unit complement(285542..285805)
/notes="putative"
/rpt_family="Alu"
misc_feature 288700..288709
/genes="TCRBV6S12"
/notes="conserved decamer; putative"
V_segment join(288811..288859,288987..289282)
/genes="TCRBV6S12"
/notes="putative"
/pseudo
exon 288811..288859
/partial
/genes="TCRBV6S12"
/notes="putative"
/pseudo
number=1
intron 288860..288986
/genes="TCRBV6S12"
/notes="putative"
number=1
exon 288987..289282
/genes="TCRBV6S12"
/notes="putative"
/pseudo
number=2
misc_signal 289283..289321
/genes="TCRBV6S12"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 289283..289289
/notes="RSS heptamer - awaiting approval of new feature
key"
misc_signal 289290..289312
/notes="RSS spacer - awaiting approval of new feature key"
misc_signal 289313..289321
/notes="RSS nonamer - awaiting approval of new feature key"
source 290393..323140
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_lines="CGM1"
/germline
/sequenced_mol="DNA"
/clone="A27"
/clone_lib="YAC 234 A6F6"
/citation=[1]
variation replace(290406..290409,"gg")
/notes="found in cosmid A27; putative"
variation replace(290776,"t")
/notes="found in cosmid A27; putative"
variation replace(292178,"g")
/notes="found in cosmid A27; putative"
variation replace(293371,"t")
/notes="found in cosmid A27; putative"
variation replace(293696,"t")
/notes="found in cosmid A27; putative"
variation replace(294162,"g")
/notes="found in cosmid A27; putative"
variation replace(294294,"g")
/notes="found in cosmid A27; putative"
variation replace(294980,"c")

```

```

repeat_unit /note="found in cosmid A27; putative"
complement(294998..295260)
/notes="putative"
/rpt_family="Alu"
variation replace(295675,"t")
/notes="found in cosmid A27; putative"
repeat_unit 295908..295979
/notes="putative"
/rpt_family="MSTA"
variation replace(295926,"c")
/notes="found in cosmid A27; putative"
variation replace(296467,"a")
/notes="found in cosmid A27; putative"
misc_feature 297155..297164
/genes="TCRBV5S3"
/notes="conserved decamer; putative"
V_segment join(297271..297319,297442..297735)
/genes="TCRBV5S3"
/notes="putative"
exon <297271..297319
/genes="TCRBV5S3"
/notes="putative"
number=1
intron 297320..297441
/genes="TCRBV5S3"
/notes="putative"
exon 297442..297735
/genes="TCRBV5S3"
/notes="putative"
number=2
misc_signal 297736..297774
/genes="TCRBV5S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 297736..297742
/notes="RSS heptamer - awaiting approval of new feature
key"
misc_signal 297743..297765
/notes="RSS spacer - awaiting approval of new feature key"
misc_signal 297766..297774
/notes="RSS nonamer - awaiting approval of new feature key"
misc_feature 302490..302499
/genes="TCRBV13S8"
/notes="conserved decamer; putative"
V_segment join(302579..302627,302717..303011)
/genes="TCRBV13S8"
/notes="putative"
exon <302579..302627
/genes="TCRBV13S8"
/notes="putative"
number=1
intron 302628..302716
/genes="TCRBV13S8"
/notes="putative"
exon 302717..303011
/genes="TCRBV13S8"
/notes="putative"
number=2
misc_signal 303012..303050
/genes="TCRBV13S8"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"

```

```

misc_signal 303012..303018
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 303019..303041
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 303042..303050
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(303862..304128)
/notes="putative"
/notes="Alu"
/notes="family="Alu"
misc_feature 306783..306792
/notes="TCRBV6S4"
/notes="conserved decamer; putative"
V_segment join(306894..306942,307089..307386)
/notes="TCRBV6S4"
/notes="putative"
/notes="putative"
exon <306894..306942
/notes="TCRBV6S4"
/notes="putative"
/notes="putative"
intron 306943..307088
/notes="TCRBV6S4"
/notes="putative"
/notes="putative"
exon 307089..307386
/notes="TCRBV6S4"
/notes="putative"
/notes="putative"
misc_signal 307387..307425
/notes="TCRBV6S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 307387..307393
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 307394..307416
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 307417..307425
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit 313325..313655
/notes="putative"
/notes="family="MSTA"
misc_feature 314667..314676
/notes="TCRBV5S2"
/notes="conserved decamer; putative"
V_segment join(314783..314831,314955..315248)
/notes="TCRBV5S2"
/notes="putative"
/notes="putative"
exon <314783..314831
/notes="TCRBV5S2"
/notes="putative"
/notes="putative"
intron 314832..314954
/notes="TCRBV5S2"
/notes="putative"
/notes="putative"
exon 314955..315248
/notes="TCRBV5S2"
/notes="putative"
/notes="putative"
misc_signal 315249..315287
/notes="TCRBV5S2"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"

```

```

misc_signal 315249..315255
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 315256..315278
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 315279..315287
/notes="RSS_nonamer - awaiting approval of new feature key"
source 321165..333651
/notes=" (vector PME15A) "
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="A2I2partial"
/clone_lib="YAC 234 A6F6"
322006..322015
/notes="TCRBV13S7"
/notes="conserved decamer; putative"
join(322095..322143,322232..322523)
/notes="partial"
/notes="TCRBV13S7"
/notes="putative"
exon <322095..322143
/notes="TCRBV13S7"
/notes="putative"
/notes="putative"
/notes="putative"
intron 322144..322231
/notes="TCRBV13S7"
/notes="putative"
exon 322232..322523
/notes="TCRBV13S7"
/notes="putative"
/notes="putative"
misc_signal 322524..322562
/notes="TCRBV13S7"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 322524..322530
/notes="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 322531..322553
/notes="RSS_spacer - awaiting approval of new feature key"
misc_signal 322554..322562
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit complement(323316..323575)
/notes="putative"
/notes="family="Alu"
misc_feature 326226..326235
/notes="TCRBV6S14"
/notes="conserved decamer; putative"
V_segment join(326337..326385,326541..326838)
/notes="TCRBV6S14"
/notes="putative"
exon <326337..326385
/notes="TCRBV6S14"
/notes="putative"
/notes="putative"
intron 326386..326540
/notes="TCRBV6S14"
/notes="putative"
/notes="putative"
exon 326541..326838
/notes="TCRBV6S14"

```

```

/misc_signal
326839..326877
/genes="TCRBV6S14"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
326839..326845
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
326846..326868
/notes="RSS_spacer - awaiting approval of new feature key"
326869..326877
/notes="RSS_nonamer - awaiting approval of new feature key"
repeat_unit
complement(332520..332789)
/notes="putative"
/rpt_family="Alu"
332724..368065
/notes="(vector PME15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone="Alf"
/clone_lib="YAC 234 A6F6"
/citation=[1]
333185..333581
/notes="putative"
/rpt_family="MSTA"
334684..334693
/genes="TCRBV5S7"
/notes="conserved decamer; putative"
join(334800..334848,334973..335266)
/genes="TCRBV5S7"
/notes="putative"
/pseudo
<334800..334848
/genes="TCRBV5S7"
/notes="putative"
/pseudo
/number=1
334849..334972
/genes="TCRBV5S7"
/notes="putative"
/number=1
334973..335266
/genes="TCRBV5S7"
/notes="putative"
/pseudo
/number=2
335267..335305
/genes="TCRBV5S7"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
335267..335273
/notes="RSS_heptamer - awaiting approval of new feature
key"
335274..335296
/notes="RSS_spacer - awaiting approval of new feature key"
335297..335305
/notes="RSS_nonamer - awaiting approval of new feature key"
342017..342026
/genes="TCRBV13S4"
/notes="conserved decamer; putative"
join(342106..342154,342243..342537)

```

```

/genes="TCRBV13S4"
/notes="putative"
<342106..342154
/genes="TCRBV13S4"
/notes="putative"
/number=1
342155..342242
/genes="TCRBV13S4"
/notes="putative"
/number=1
342243..342537
/genes="TCRBV13S4"
/number=2
342538..342576
/genes="TCRBV13S4"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
342538..342544
/notes="RSS_heptamer - awaiting approval of new feature
key"
342545..342567
/notes="RSS_spacer - awaiting approval of new feature key"
342568..342576
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(343376..343641)
/notes="putative"
/rpt_family="Alu"
346609..346618
/genes="TCRBV6S3"
/notes="conserved decamer; putative"
<346720..346768
/genes="TCRBV6S3"
/notes="putative"
/number=1
join(346720..346768,346906..347203)
/genes="TCRBV6S3"
/notes="putative"
346769..346905
/genes="TCRBV6S3"
/notes="putative"
/number=1
346906..347203
/genes="TCRBV6S3"
/notes="putative"
/number=2
347204..347242
/genes="TCRBV6S3"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
347204..347210
/notes="RSS_heptamer - awaiting approval of new feature
key"
347211..347233
/notes="RSS_spacer - awaiting approval of new feature key"
347234..347242
/notes="RSS_nonamer - awaiting approval of new feature key"
349927..350188
/notes="putative"
/rpt_family="Alu"
354820..354829
/genes="TCRBV5S8"
/notes="conserved decamer; putative"
join(354936..354984,355109..355402)
/genes="TCRBV5S8"

```

```

exon      /note="putative"
<354936..354984
/feature="TCRBV538"
/feature="putative"
/number=1
intron    /gene="TCRBV538"
354985..355108
/feature="putative"
/number=1
exon      355109..355402
/feature="TCRBV538"
/feature="putative"
/number=2
misc_signal 355403..355441
/feature="TCRBV538"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 355403..355409
/feature="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 355410..355432
/feature="RSS_spacer - awaiting approval of new feature key"
misc_signal 355433..355441
/feature="RSS_nonamer - awaiting approval of new feature key"
misc_feature 364026..364035
/feature="TCRBV655"
/feature="conserved decamer; putative"
V_segment join(364137..364185,364312..364609)
/feature="TCRBV655"
/feature="putative"
exon      <364137..364185
/feature="TCRBV655"
/feature="putative"
/number=1
intron    364186..364311
/feature="TCRBV655"
/feature="putative"
/number=1
exon      364312..364609
/feature="TCRBV655"
/feature="putative"
/number=2
source    364352..402533
/organism="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="H7.1"
/citation=[2]
misc_signal 364610..364648
/feature="TCRBV655"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 364610..364648
/feature="TCRBV655"
/feature="RSS" - awaiting approval of new feature key;
heptamer is at 5' end; putative"
misc_signal 364610..364616
/feature="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 364610..364616
/feature="RSS_heptamer - awaiting approval of new feature
key"
misc_signal 364617..364639
/feature="RSS_spacer - awaiting approval of new feature key"

```

```

misc_signal 364617..364639
/feature="RSS_spacer - awaiting approval of new feature key"
364640..364648
/feature="RSC_nonamer - awaiting approval of new feature key"
364640..364648
/feature="RSC_nonamer - awaiting approval of new feature key"
variation replace(364925,"c")
/feature="found in cosmid H7.1; putative"
variation replace(365416,"g")
/feature="found in cosmid H7.1; putative"
variation replace(365931,"t")
/feature="found in cosmid H7.1; putative"
variation replace(366037..366038,"att")
/feature="found in cosmid H7.1; putative"
variation replace(366208,"t")
/feature="found in cosmid H7.1; putative"
variation replace(366212,"g")
/feature="found in cosmid H7.1; putative"
variation replace(366343,"g")
/feature="found in cosmid H7.1; putative"
variation replace(366405,"g")
/feature="found in cosmid H7.1; putative"
variation replace(366490..366491,"tgg")
/feature="found in cosmid H7.1; putative"
variation replace(366568..366573,"tt")
/feature="found in cosmid H7.1; putative"
variation replace(366621,"c")
/feature="found in cosmid H7.1; putative"
variation replace(366773..366774,"ata")
/feature="found in cosmid H7.1; putative"
variation replace(366845,"t")
/feature="found in cosmid H7.1; putative"
variation replace(367512,"a")
/feature="found in cosmid H7.1; putative"
variation replace(367521,"t")
/feature="found in cosmid H7.1; putative"
variation replace(367635,"a")
/feature="found in cosmid H7.1; putative"
variation replace(367830..367831,"att")
/feature="found in cosmid H7.1; putative"
repeat_unit complement(367842..368114)
/feature="putative"
/feature="Alu"
variation replace(367844,"c")
/feature="found in cosmid H7.1; putative"
variation replace(367895,"t")
/feature="found in cosmid H7.1; putative"
variation replace(367952,"c")
/feature="found in cosmid H7.1; putative"
variation replace(368035,"g")
/feature="found in cosmid H7.1; putative"
variation replace(368065,"c")
/feature="found in cosmid H7.1; putative"
misc_feature 370564..370573
/feature="TCRBV23S1"
/feature="conserved decamer; putative"
V_segment join(370656..370734,370845..371139)
/feature="TCRBV23S1"
/feature="putative"
exon      <370656..370734
/feature="TCRBV23S1"
/feature="putative"
/number=1
intron    370735..370844

```

```

/ gene="TCRBV23S1"
/ note="putative"
/ number=1
370845..371139
/ gene="TCRBV23S1"
/ note="putative"
/ number=2
371140..371178
/ gene="TCRBV23S1"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
371140..371146
/ note="RSS_heptamer - awaiting approval of new feature
key"
371147..371169
/ note="RSS_spacer - awaiting approval of new feature key"
371170..371178
/ note="RSS nonamer - awaiting approval of new feature key"
complement(371174..371631)
/ note="putative"
/ rpt_family="MLT1B"
complement(371180..372141)
/ note="putative"
/ rpt_family="Alu"
complement(373018..373274)
/ note="putative"
/ rpt_family="Alu"
375722..375975
/ note="putative"
/ rpt_family="Alu"
complement(377891..378378)
/ note="putative"
/ rpt_family="MLT2CA"
378984..378993
/ gene="TCRBV12S2"
/ note="conserved decamer; putative"
join(379068..379116,379223..379517)
/ gene="TCRBV12S2"
/ note="putative"
<379068..379116
/ gene="TCRBV12S2"
/ note="putative"
/ number=1
379117..379222
/ gene="TCRBV12S2"
/ note="putative"
/ number=1
379223..379517
/ gene="TCRBV12S2"
/ note="putative"
/ number=2
379518..379556
/ gene="TCRBV12S2"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
379518..379524
/ note="RSS_heptamer - awaiting approval of new feature
key"
379525..379547
/ note="RSS_spacer - awaiting approval of new feature key"
379548..379556
/ note="RSS nonamer - awaiting approval of new feature key"
380360..380626
/ note="putative"

```

```

/ rpt_family="Alu"
380936..381190
/ note="putative"
/ rpt_family="Alu"
381368..420906
/ organism="Homo sapiens"
/ germline
/ sequenced_mol="DNA"
/ clone="H12.18"
/ citations=[2]
382199..382460
/ note="putative"
/ rpt_family="Alu"
382674..383012
/ note="putative"
/ rpt_family="MSTA"
complement(383847..384274)
/ note="putative"
/ rpt_family="LINE 1"
384754..385422
/ note="putative"
/ rpt_family="LINE 1"
389605..389615
/ gene="TCRBV21S4"
/ note="conserved decamer; putative"
join(389716..389764,389856..390153)
/ gene="TCRBV21S4"
/ note="putative"
<389716..389764
/ gene="TCRBV21S4"
/ note="putative"
/ number=1
389765..389855
/ gene="TCRBV21S4"
/ note="putative"
/ number=1
389856..390153
/ gene="TCRBV21S4"
/ note="putative"
/ number=2
390154..390192
/ gene="TCRBV21S4"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
390154..390160
/ note="RSS_heptamer - awaiting approval of new feature
key"
390161..390183
/ note="RSS_spacer - awaiting approval of new feature key"
390184..390192
/ note="RSS nonamer - awaiting approval of new feature key"
complement(393711..394157)
/ note="putative"
/ rpt_family="MLT1D"
395203..395212
/ gene="TCRBV8S1"
/ note="conserved decamer; putative"
join(395318..395366,395467..395764)
/ gene="TCRBV8S1"
/ note="putative"
<395318..395366
/ gene="TCRBV8S1"
/ note="putative"
/ number=1

```

```

intron      395367..395466
            /gene="TCRBV8S1"
            /note="putative"
            /number=1
exon        395467..395764
            /gene="TCRBV8S1"
            /note="putative"
            /number=2
misc_signal 395765..395803
            /gene="TCRBV8S1"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            395765..395771
            /note="RSS_heptamer - awaiting approval of new feature
            key"
            395772..395794
            /note="RSS_spacer - awaiting approval of new feature key"
            395795..395803
            /note="RSS_nonamer - awaiting approval of new feature key"
            complement(397023..397460)
            /note="putative"
            /rpt_family="MLTID"
            398526..398535
            /gene="TCRBV8S2"
            /note="conserved decamer; putative"
            join(398641..398689,398790..399087)
            /gene="TCRBV8S2"
            /note="putative"
            <398641..398689
            /gene="TCRBV8S2"
            /note="putative"
            /number=1
            398690..398789
            /gene="TCRBV8S2"
            /note="putative"
            /number=1
            398790..399087
            /gene="TCRBV8S2"
            /note="putative"
            /number=2
            399088..399126
            /gene="TCRBV8S2"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            399088..399094
            /note="RSS_heptamer - awaiting approval of new feature
            key"
            399095..399117
            /note="RSS_spacer - awaiting approval of new feature key"
            399118..399126
            /note="RSS_nonamer - awaiting approval of new feature key"
            402029..402832
            /note="putative"
            /rpt_family="HUMERSP2A"
            complement(403915..404184)
            /note="putative"
            404140..442094
            /rpt_family="Alu"
            /organism="Homo sapiens"
            /germline
            /sequenced_mol="DNA"
            /clone="H130"
            /citation={2}
            405472..409693
            repeat_unit

```

```

repeat_unit 409045..409306
            /note="putative"
            /rpt_family="Alu"
            complement(409793..410042)
            /note="putative"
            /rpt_family="Alu"
            410544..411352
            /note="putative"
            /rpt_family="HUMERVKB1"
            411803..412066
            /note="putative"
            /rpt_family="Alu"
            412870..413127
            /note="putative"
            /rpt_family="Alu"
            415714..415723
            /gene="TCRBV8S3"
            /note="conserved decamer; putative"
            join(415832..415880,415981..416278)
            /gene="TCRBV8S3"
            /note="putative"
            <415832..415880
            /gene="TCRBV8S3"
            /note="putative"
            /number=1
            415881..415980
            /gene="TCRBV8S3"
            /note="putative"
            /number=1
            415981..416278
            /gene="TCRBV8S3"
            /note="putative"
            /number=2
            416279..416317
            /gene="TCRBV8S3"
            /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            416279..416285
            /note="RSS_heptamer - awaiting approval of new feature
            key"
            416286..416308
            /note="RSS_spacer - awaiting approval of new feature key"
            416309..416317
            /note="RSS_nonamer - awaiting approval of new feature key"
            420729..421157
            /note="putative"
            /rpt_family="MLTID"
            422664..422673
            /gene="TCRBV16S1"
            /note="conserved decamer; putative"
            join(422775..422823,422910..423207)
            /gene="TCRBV16S1"
            /note="putative"
            <422775..422823
            /gene="TCRBV16S1"
            /note="putative"
            /number=1
            422824..422909
            /gene="TCRBV16S1"
            /note="putative"
            /number=1
            422910..423207
            exon

```

```

/misc_signal      /gene="TCRBV16S1"
                  /note="putative"
                  /number=2
423208..423246
/gene="TCRBV16S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal      423208..423214
/note="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal      423215..423237
/note="RSS_spacer - awaiting approval of new feature key"
/misc_signal      423238..423246
/note="RSS_nonamer - awaiting approval of new feature key"
repeat_unit      complement(4232957..424217)
/note="putative"
/misc_feature      /rpt_family="Alu"
427725..427734
/gene="TCRBV24S1"
/note="conserved decamer; putative"
V_segment         join(427850..427898,428024..428318)
/gene="TCRBV24S1"
/note="putative"
exon              <427850..427898
/gene="TCRBV24S1"
/note="putative"
intron            /number=1
427899..428023
/gene="TCRBV24S1"
/note="putative"
exon              428024..428318
/gene="TCRBV24S1"
/note="putative"
/misc_signal      428319..428357
/gene="TCRBV24S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal      428319..428325
/note="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal      428326..428348
/note="RSS_spacer - awaiting approval of new feature key"
/misc_signal      428349..428357
/note="RSS_nonamer - awaiting approval of new feature key"
/misc_feature      432763..432772
/note="conserved decamer; putative"
V_segment         join(432859..432907,433015..433312)
/partial
/gene="TCRBV25S1"
/note="putative"
exon              /pseudo
<432859..432907
/gene="TCRBV25S1"
/note="putative"
intron            /pseudo
432908..433014
/gene="TCRBV25S1"
/note="putative"
exon              /number=1
433015..433312
/gene="TCRBV25S1"

```

```

/misc_signal      /note="putative"
                  /pseudo
                  /number=2
433313..433351
/gene="TCRBV25S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal      433313..433319
/gene="TCRBV25S1"
/note="RSS_heptamer - awaiting approval of new feature
key; putative"
/misc_signal      433320..433342
/gene="TCRBV25S1"
/note="RSS_spacer - awaiting approval of new feature key;
putative"
/misc_signal      433343..433351
/note="RSS_nonamer - awaiting approval of new feature key"
/misc_signal      433343..433349
/note="RSS_nonamer - awaiting approval of new feature key;
putative"
repeat_unit       433523..433794
/note="putative"
repeat_unit       /rpt_family="Alu"
434353..434613
/note="putative"
repeat_unit       /rpt_family="Alu"
435607..435864
/note="putative"
/misc_feature      /rpt_family="Alu"
436423..436433
/gene="TCRBV26S1"
/note="conserved decamer; putative"
V_segment         join(436475..436523,436915..437209)
/gene="TCRBV26S1"
/note="putative"
exon              /pseudo
/codon_start=1
<436475..436523
/gene="TCRBV26S1"
/note="putative"
intron            /pseudo
436524..436914
/gene="TCRBV26S1"
/note="putative"
repeat_unit       /number=1
436572..436830
/note="putative"
exon              /rpt_family="Alu"
436915..437209
/gene="TCRBV26S1"
/note="putative"
/misc_signal      /pseudo
437210..437248
/gene="TCRBV26S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal      437210..437216
/note="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal      437217..437239
/note="RSS_spacer - awaiting approval of new feature key"
/misc_signal      437240..437248

```

/note="RSS nonamer - awaiting approval of new feature key"

```

source
438704..472940
/note="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="Alf6"
/clone_lib="YAC 234 A6F6"
/citation=[1]
replace(439095,"a")
/note="from cosmid A16; putative"
replace(439470..439471,"cctt")
/note="from cosmid A16; putative"
replace(439591,"t")
/note="from cosmid A16; putative"
440365..440622
/note="putative"
/rpt_family="Alu"
replace(440424,"c")
/note="from cosmid A16; putative"
replace(440641..440642,"caataata")
/note="from cosmid A16; putative"
replace(440936..440939,"agtg")
/note="from cosmid A16; putative"
replace(441417,"c")
/note="from cosmid A16; putative"
replace(441798..441799,"taa")
/note="from cosmid A16; putative"
441863..442125
/note="putative"
/rpt_family="Alu"
442182..442207
/note="polymorphic"
/rpt_family="microsatellite"
/rpt_unit=442182..442183
/evidence=experimental
complement(442683..442948)
/note="putative"
/rpt_family="Alu"
444936..444983
/note="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=444936..444937
448412..448665
/note="putative"
/rpt_family="Alu"
449251..449509
/note="putative"
/rpt_family="Alu"
449974..450226
/note="putative"
/rpt_family="Alu"
450430..450439
/note="TCRBV18S1"
/note="conserved decamer; putative"
Join(450546..450594,450867..451164)
/note="TCRBV18S1"
/note="heptamer is at 5' end; putative"
<450546..450594
/note="TCRBV18S1"
/note="putative"
/number=1
450595..450866

```

```

/genes="TCRBV18S1"
/note="putative"
/number=1
450867..451164
/genes="TCRBV18S1"
/note="putative"
/number=2
451165..451203
/genes="TCRBV18S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
451165..451171
/note="RSS heptamer - awaiting approval of new feature
key"
451172..451194
/note="RSS spacer - awaiting approval of new feature key"
451195..451203
/note="RSS nonamer - awaiting approval of new feature key"
452324..486943
/note="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="C215"
/clone_lib="YAC 234 A72B3"
/citation=[1]
Join(453806..453854,453987..454281)
/genes="TCRBV17S1"
/note="putative"
<453806..453854
/genes="TCRBV17S1"
/note="putative"
/number=1
453987..454281
/genes="TCRBV17S1"
/note="putative"
/number=2
454282..454320
/genes="TCRBV17S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
454282..454288
/note="RSS heptamer - awaiting approval of new feature
key"
454289..454311
/note="RSS spacer - awaiting approval of new feature key"
454312..454320
/note="RSS nonamer - awaiting approval of new feature key"
complement(454622..454883)
/note="putative"
/rpt_family="Alu"
457907..>457907
/note="translocation breakpoint; t(7:9); putative"
complement(461020..461281)
/note="putative"
/rpt_family="Alu"
Join(461476..461506,461845..462148)
/genes="TCRBV2S1"
/note="putative"

```

```

exon
<461476..461506
/ gene="TCRBV251"
/ note="putative"
/ number=1
461507..461844
/ gene="TCRBV251"
/ note="putative"
/ number=1
461845..462148
/ gene="TCRBV251"
/ number=2
462149..462187
/ gene="TCRBV251"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
462149..462155
/ gene="TCRBV251"
/ note="RSS heptamer - awaiting approval of new feature
key; putative"
462156..462178
/ gene="TCRBV251"
/ note="RSS spacer - awaiting approval of new feature key;
putative"
462179..462187
/ gene="TCRBV251"
/ note="RSS nonamer - awaiting approval of new feature key;
putative"
465204..465245
/ note="polymorphic; putative"
/ rpt_family="microsatellite"
/ rpt_unit=465204..465205
468059..468314
/ note="putative"
/ rpt_family="Alu"
468548..468804
/ note="putative"
/ rpt_family="Alu"
470316..470574
/ note="putative"
/ rpt_family="Alu"
470593..470624
/ note="polymorphic; putative"
/ rpt_family="microsatellite"
/ rpt_unit=470593..470596
470746..471356
/ note="putative"
/ rpt_family="LINE 1"
471558..471567
/ gene="TCRBV10S1"
/ note="conserved decamer; putative"
join(471659..471708,471822..472119)
/ gene="TCRBV10S1"
/ note="putative"
/ pseudo
471659..471708
/ partial
/ gene="TCRBV10S1"
/ note="putative"
/ pseudo
471709..471821
/ gene="TCRBV10S1"
/ note="putative"
/ number=1
471709..471821
/ gene="TCRBV10S1"
/ note="putative"
/ number=1

```

```

exon
471822..472119
/ gene="TCRBV10S1"
/ note="putative"
/ pseudo
/ number=2
472120..472158
/ gene="TCRBV10S1"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
472120..472126
/ note="RSS heptamer - awaiting approval of new feature
key"
472127..472149
/ note="RSS spacer - awaiting approval of new feature key"
472150..472158
/ note="RSS nonamer - awaiting approval of new feature key"
complement(472732..472994)
/ note="putative"
/ rpt_family="Alu"
complement(475772..475851)
/ note="putative"
/ rpt_family="MIR"
476399..476408
/ gene="TCRBV29S1"
/ note="conserved decamer; putative"
join(476480..476528,476661..476730)
/ gene="TCRBV29S1"
/ note="putative"
/ pseudo
<476480..476528
/ gene="TCRBV29S1"
/ note="putative"
/ pseudo
/ number=1
476529..476660
/ gene="TCRBV29S1"
/ note="putative"
/ number=1
476661..476730
/ gene="TCRBV29S1"
/ note="putative"
/ pseudo
/ number=2
476931..476969
/ gene="TCRBV29S1"
/ note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
476931..476937
/ note="RSS heptamer - awaiting approval of new feature
key"
476938..476960
/ note="RSS spacer - awaiting approval of new feature key"
476961..476969
/ note="RSS nonamer - awaiting approval of new feature key"
480626..480635
/ gene="TCRBV19S1"
/ note="conserved decamer; putative"
join(480699..480747,480898..481195)
/ gene="TCRBV19S1"
/ note="putative"
/ pseudo
<480699..480747
/ gene="TCRBV19S1"
/ note="putative"

```

```

/misc_signal
/number=1
480748..480897
/genes="TCRBV19S1"
/notes="lacks functional splice donor; putative; does not
fit consensus"
/number=1
480898..481195
/genes="TCRBV19S1"
/notes="putative"
/pseudo
/number=2
481196..481234
/genes="TCRBV19S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
481196..481202
/notes="RSS_heptamer - awaiting approval of new feature
key"
/misc_signal
481203..481225
/notes="RSS_spacer - awaiting approval of new feature key"
/misc_signal
481226..481234
/notes="RSS_nonamer - awaiting approval of new feature key"
source
482597..515832
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="ATCC 1521"
/germline
/sequenced_mol="DNA"
/clones="G1"
/clone_lib="Kai Wang's"
/citation=[1]
replace(484618,"c")
/notes="from cosmid G1; putative"
485079..487500
/notes="putative"
/rpt_family="LINE 1"
488729..489015
/notes="putative"
/rpt_family="HUMERSP2A"
489941..489950
/genes="TCRBV15S1"
/notes="conserved decamer; putative"
Join(490040..490088,490221..490516)
/genes="TCRBV15S1"
/notes="putative"
<490040..490088
/genes="TCRBV15S1"
/notes="putative"
/number=1
490089..490220
/genes="TCRBV15S1"
/notes="putative"
/number=1
490221..490516
/genes="TCRBV15S1"
/notes="putative"
/number=2
490517..490555
/genes="TCRBV15S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
/misc_signal
490517..490523
/notes="RSS_heptamer - awaiting approval of new feature

```

```

key"
490524..490546
/notes="RSS_spacer - awaiting approval of new feature key"
490547..490555
/notes="RSS_nonamer - awaiting approval of new feature key"
complement(493122..493784)
/notes="putative"
/rpt_family="LINE 1"
complement(493240..493514)
/notes="putative"
/rpt_family="Alu"
complement(494073..495301)
/notes="putative"
/rpt_family="LINE 1"
complement(497524..498018)
/notes="putative"
/rpt_family="MER2"
501895..502102
/notes="putative"
504326..504335
/genes="TCRBV11S1"
/notes="conserved decamer; putative"
Join(504415..504463,504588..504882)
/genes="TCRBV11S1"
/notes="putative"
504415..504463
/partial
/genes="TCRBV11S1"
/notes="putative"
/number=1
504464..504587
/genes="TCRBV11S1"
/notes="putative"
/number=1
504588..504882
/genes="TCRBV11S1"
/notes="putative"
/number=2
504707..538564
/notes="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="CGM1"
/germline
/sequenced_mol="DNA"
/clone_lib="YAC 199 B157G9"
/citation=[1]
504883..504921
/genes="TCRBV11S1"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
504883..504889
/notes="RSS_heptamer - awaiting approval of new feature
key"
504890..504912
/notes="RSS_spacer - awaiting approval of new feature key"
504913..504921
/notes="RSS_nonamer - awaiting approval of new feature key"
506376..506604
/notes="putative"
/rpt_family="MER2"
507025..507859
/notes="putative"

```

```

repeat_unit      /rpt_family="LINE 1"
                  complement(508000..510053)
                  /note="putative"
variation         /rpt_family="LINE 1"
                  replace(508199,"g")
                  /note="from cosmid C68; putative"
repeat_unit      508498..508759
                  /note="putative"
variation         /rpt_family="Alu"
                  replace(510841..510843,"ct")
                  /note="from cosmid C68; putative"
repeat_unit      complement(510867..511137)
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  512915..514739
                  /note="putative"
repeat_unit      /rpt_family="LINE 1"
                  513363..513884
                  /note="putative"
variation         /rpt_family="MLT2A"
                  replace(513888,"c")
                  /note="from cosmid C68; putative"
V_segment        join(515044..515090,515129..515497)
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
exon              <515044..515090
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
intron            515091..515128
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
exon              515129..515497
                  /gene="TCRBV33S1"
                  /note="putative"
                  /pseudo
misc_signal       /codon_start=1
                  515498..515536
                  /gene="TCRBV33S1"
                  /note="RSS - awaiting approval of new feature key;
                  heptamer is at 5' end; putative"
misc_signal       /note="RSS_heptamer - awaiting approval of new feature
                  key"
misc_signal       515505..515527
                  /note="RSS_spacer - awaiting approval of new feature key"
misc_signal       515528..515536
                  /note="RSS_nonamer - awaiting approval of new feature key"
repeat_unit      518289..518549
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  complement(520882..521434)
                  /note="putative"
repeat_unit      /rpt_family="LINE 1"
                  523833..524090
                  /note="putative"
repeat_unit      /rpt_family="Alu"
                  524764..525025
                  /note="putative"

```

```

V_segment        /rpt_family="Alu"
                  join(529336..529384,529526..529820)
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
exon              <529336..529384
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
                  /number=1
intron            /codon_start=1
                  529385..529525
                  /gene="TCRBV28S1"
                  /note="putative"
                  /number=1
exon              529526..529820
                  /gene="TCRBV28S1"
                  /note="putative"
                  /pseudo
                  /number=2
misc_signal       /codon_start=1
                  529821..529859
                  /gene="TCRBV28S1"
                  /note="RSS - awaiting approval of new feature key;
                  heptamer is at 5' end; putative"
                  529821..529827
                  /gene="TCRBV28S1"
                  /note="RSS_heptamer - awaiting approval of new feature
                  key; putative"
                  529828..529850
                  /gene="TCRBV28S1"
                  /note="RSS_spacer - awaiting approval of new feature key;
                  putative"
                  529851..529859
                  /gene="TCRBV28S1"
                  /note="RSS_nonamer - awaiting approval of new feature key;
                  putative"
                  530024..530166
                  /note="putative"
                  /rpt_family="MER4"
                  complement(530975..531036)
                  /note="putative"
                  /rpt_family="MLT1A"
                  531638..565300
                  /note="(vector PWE15A)"
                  /organism="Homo sapiens"
                  /cell_line="CGM1"
                  /germline
                  /sequenced_mol="DNA"
                  /clone="C21"
                  /clone_lib="YAC 234 A72B3"
                  /citation=[1]
                  replace(531984..531985,"gcc")
                  /note="found in cosmid C21; putative"
                  replace(533511..533512,"tgg")
                  /note="found in cosmid C21; putative"
                  533983..534356
                  /note="putative"
                  /rpt_family="THE transposon like element"
                  535432..535690
                  /note="putative"
                  /rpt_family="Alu"
                  536867..537285

```

```

/note="putative"
/rpt_family="MLTID"
replace(537605..537607,"ta")
/note="found in cosmid C21; putative"
539683..539720
/note="polymorphic; putative"
/rpt_family="microsatellite"
/rpt_unit=539683..539684
544636..544866
/note="putative"
/rpt_family="MLTIB"
join(545039..545080,545214..545572)
/genes="TCRBV34S1"
/note="putative"
/pseudo
/codon_start=1
<545039..545080
/genes="TCRBV34S1"
/note="putative"
/pseudo
/number=1
/codon_start=1
545081..545213
/genes="TCRBV34S1"
/note="putative"
/number=1
545214..545572
/genes="TCRBV34S1"
/note="putative"
/pseudo
/number=2
/codon_start=1
545573..545579
/genes="TCRBV34S1"
/note="RSS - awaiting approval of new feature key;
degraded RSS; putative"
545573..545579
/genes="TCRBV34S1"
/note="RSS heptamer - awaiting approval of new feature
key; putative"
complement(545711..546155)
/note="putative"
/rpt_family="MLTID"
548386..548634
/note="putative"
/rpt_family="Alu"
548863..548872
/genes="TCRBV14S1"
/note="conserved decamer; putative"
join(549038..549086,549216..549510)
/genes="TCRBV14S1"
/note="putative"
<549038..549086
/genes="TCRBV14S1"
/note="putative"
/number=1
549087..549215
/genes="TCRBV14S1"
/note="putative"
/number=1
549216..549510
/genes="TCRBV14S1"
/note="putative"
/number=2

```

```

misc_signal 549511..549549
/genes="TCRBV14S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
549511..549549
/genes="TCRBV14S1"
/note="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
549511..549517
/note="RSS heptamer - awaiting approval of new feature
key"
549518..549540
/note="RSS spacer - awaiting approval of new feature key"
549541..549549
/note="RSS nonamer - awaiting approval of new feature key"
551440..551700
/note="putative"
/rpt_family="Alu"
552491..590912
/note="(vector PWE15A)"
/organism="Homo sapiens"
/cell_line="ATCC 1521"
/germline
/sequenced_mol="DNA"
/clone="X11"
/clone_lib="Kai Wang's"
/citation=[1]
553554..553886
/note="putative"
/rpt_family="MLTIB"
554259..554268
/genes="TCRBV3S1"
/note="conserved decamer; putative"
join(554328..554376,554514..554808)
/partial
/genes="TCRBV3S1"
/note="This CDS feature is included to show the
translation of the corresponding V segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624710"
/codon_start=1
/translations="MCIBLLCRVAFCLAVGLVDVKVQTQSSRYLVKRTCEKVFLECVQ
DMDHENWVRQDPGLGLRLIYFSYDVVKKEKGDIPGYSVSREKKEKRFSLIESAST
NQTSMYLCASSL"
join(554328..554376,554514..554808)
/genes="TCRBV3S1"
/note="putative"
<554328..554376
/genes="TCRBV3S1"
/note="putative"
/number=1
554377..554513
/genes="TCRBV3S1"
/note="putative"
/number=1
replace(554495,"a")
/note="found in cosmid X11; putative"
554514..554808
/genes="TCRBV3S1"
/note="putative"
/number=2
554809..554847
/genes="TCRBV3S1"

```

```

misc_signal      /note="RSS - awaiting approval of new feature key";
                  heptamer is at 5' end; putative"
554809..554815
/note="RSS_heptamer - awaiting approval of new feature
key"
554816..554838
/note="RSS_spacer - awaiting approval of new feature key"
554839..554847
/note="RSS_nonamer - awaiting approval of new feature key"
complement(557145..557400)
/note="putative"
/rpt family="Alu"
557655..558216
/note="putative"
/rpt family="LINE1"
559309..561579
/note="putative"
/rpt family="THE1 transposon like element"
replace(561498,"a")
/note="found in cosmid X11; putative"
replace(561742,"g")
/note="found in cosmid X11; putative"
replace(562192,"c")
/note="from cosmid X11; putative"
replace(562232,"a")
/note="from cosmid X11; putative"
replace(562694,"a")
/note="from cosmid X11; putative"
replace(563416,"c")
/note="from cosmid X11; putative"
replace(563871,"t")
/note="from cosmid X11; putative"
replace(564119,"t")
/note="from cosmid X11; putative"
565220
/note="The 3' breakpoint occurs at basepair 592356,
602910, 613477, or 623282 -- identical positions within a
homology unit that is tandemly repeated.; translocation
breakpoint t(7:9); putative"
567741..568016
/note="putative"
/rpt family="Alu"
567946..569865
/note="putative"
/rpt family="LINE 1"
568023..569813
/note="putative"
/rpt family="LINE1"
569393..606124
/note="(vector PME15A)"
/orqanlem="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="X6a"
/clone_lib="Kai Wang's"
/tissue_lib="ATCC 1521"
/citation=[1]
replace(570855,"c")
/note="from cosmid X6A; putative"
replace(570966..570968,"aa")
/note="from cosmid X6A; putative"
replace(572279,"a")
/note="from cosmid X6A; putative"
replace(573140,"a")

```

```

variation        /note="from cosmid X6A; putative"
                  replace(573457,"a")
/note="from cosmid X6A; putative"
replace(573462,"a")
/note="from cosmid X6A; putative"
join(573946..573979,574259..574559)
/genes="TCRBV4S1"
/note="putative"
/codon_start=1
<573946..573979
/genes="TCRBV4S1"
/note="putative"
/number=1
join(573946..573979,574259..574559)
/partial
/genes="TCRBV4S1"
/note="This CDS feature is included to show the
translation of the corresponding V_segment. Presently
translation qualifiers on V_segment features are illegal.
NCBI gi: 624711"
/codon_start=1
/translation="MISLLILLGLGVFSAVISQPSRDICQGTSLTICQVDQSV
TMMFWRQPCQSLTIATANQSGSEATYESGFVDKPFISRPNLTFSTLTYSNNSPD
SSIIYLCSE"
573980..574258
/genes="TCRBV4S1"
/note="putative"
/number=1
replace(574024..574025,"gggtggaag")
/note="from cosmid X6A; putative"
574259..574559
/genes="TCRBV4S1"
/note="putative"
/number=2
574560..574566
/note="RSS_heptamer - awaiting approval of new feature
key"
574567..574589
/note="RSS_spacer - awaiting approval of new feature key"
574590..574598
/note="RSS_nonamer - awaiting approval of new feature key"
replace(575119,"c")
/note="from cosmid X6A; putative"
replace(578315..578317,"ca")
/note="from cosmid X6A; putative"
replace(578641,"c")
/note="from cosmid X6A; putative"
replace(579597,"t")
/note="from cosmid X6A; putative"
replace(579607,"c")
/note="from cosmid X6A; putative"
replace(579794,"c")
/note="from cosmid X6A; putative"
replace(579842..579843,"aaga")
/note="from cosmid X6A; putative"
580170..580273
/note="putative"
/rpt family="MIR1"
complement(580171..580256)
/note="putative"

```

```

variation      /rpt_family="MIR"
repeat_unit    replace(580374,"g")
               /note="from cosmid X6A; putative"
               580376..580477
               /note="putative"
               /rpt_family="MIR"
variation      replace(580467,"g")
               /note="from cosmid X6A; putative"
variation      replace(580471,"ca")
               /note="from cosmid X6A; putative"
variation      replace(581770,"a")
               /note="from cosmid X6A; putative"
variation      replace(582745,"t")
               /note="from cosmid X6A; putative"
variation      replace(582949,"a")
               /note="from cosmid X6A; putative"
variation      replace(583192,584223..584382,585442..585695,
mRNA           586099..586235,586536..586688)
               /gene="TRYA"
exon           <583153..583192
               /gene="TRYA"
               /note="putative"
               /number=1
CDS            join(583153..583192,584223..584382,585442..585695,
               586099..586235,586536..586688)
               /gene="TRYA"
               /note="putative; NCBI gi: 624712"
               /codon_start=1
               /product="trypsinogen A"
               /translation="MNPILLITFVAALAAPFDDDDKIVGCVNCENSVPYQVSLNSG
               YHF CGGLINEQVMVSGHCYKSRIVRIGEHNIIEVLENEQF INAAKTIIRHPQYDRK
               TLNIDIMLKLSRAVNRVTSISLTAPPATGTKCLISGMGTASSGADYPDELQC
               LDAPVLQAKCEASVPKITSNMFVGFLEGCKSCQGDGPGPVVNCGLQGVNSWGD
               GCAQRKPGVYTKVNTYVVRWIKNTIAANS"
               583193..584222
               /gene="TRYA"
               /note="putative; does not fit consensus"
               /number=1
variation      replace(583491,"a")
exon           /note="from cosmid X6A; putative"
               584223..584382
               /gene="TRYA"
               /number=2
intron         584383..585441
               /gene="TRYA"
               /note="putative"
               /number=2
variation      replace(585107,"c")
exon           /note="from cosmid X6A; putative"
               585442..585695
               /gene="TRYA"
               /note="putative"
               /number=3
intron         585696..586098
               /gene="TRYA"
               /note="putative"
               /number=3
exon           586099..586235
               /gene="TRYA"
               /note="putative"
               /number=4
variation      replace(586130,"t")
               /gene="TRYA"
               /note="from cosmid X6A; putative"

```

```

intron         586236..586535
               /gene="TRYA"
               /note="putative"
               /number=4
exon           586536..586688
               /gene="TRYA"
               /note="putative"
               /number=5
variation      replace(586682,"t")
               /gene="TRYA"
               /note="from cosmid X6A; putative"
               586719..586724
               /note="putative"
polyA_signal   replace(587095,"t")
               /note="from cosmid X6A; putative"
variation      replace(587353,"g")
               /note="from cosmid X6A; putative"
variation      replace(587365,"a")
               /note="from cosmid X6A; putative"
variation      replace(588389,"a")
               /note="from cosmid X6A; putative"
               588490..588490
               /note="from cosmid X6A; putative"
               588898..588898
               /note="from cosmid X6A; putative"
               complement(589057..589313)
               /note="putative"
               /rpt_family="Alu"
               replace(589124,"c")
               /note="from cosmid X6A; putative"
               589531..590009
               /note="putative"
               /rpt_family="LINE1"
               replace(589535,"c")
               /note="from cosmid X6A; putative"
               replace(589738,"c")
               /note="from cosmid X6A; putative"
               replace(589740,"t")
               /note="from cosmid X6A; putative"
               replace(590011..590015,at")
               /note="from cosmid X6A; putative"
               replace(590085,"g")
               /note="from cosmid X6A; putative"
               590218..590472
               /note="putative"
               /rpt_family="Alu"
               590287..590435
               /note="putative"
               /rpt_family="Alu"
               replace(590308,"g")
               /note="from cosmid X6A; putative"
               replace(590334,"c")
               /note="from cosmid X6A; putative"
               replace(590506,"a")
               /note="from cosmid X6A; putative"
               replace(590509..590510,atg")
               /note="from cosmid X6A; putative"
               replace(591483,"t")
               /note="from cosmid X6A; putative"
               replace(592143,"a")
               /note="from cosmid X6A; putative"
               replace(592154,"a")
               /note="from cosmid X6A; putative"
               join(<594083..594122,595142..595301,596361..596614,
mRNA

```

```

597022..597158,597460..597610)
/feature="TRYB"
/feature="putative"
<594083..594122
/feature="TRYB"
/feature="putative"
/pseudo
/number=1
Join(594083..594122,595142..595301,596361..596614,
597022..597158,597460..597610)
/feature="TRYB"
/feature="putative"
/pseudo
/product="trypsinogen B"
594123..595141
/partial
/feature="TRYB"
/feature="putative"
/number=1
595142..595301
/feature="TRYB"
/feature="putative"
/pseudo
/number=2
595302..596360
/feature="TRYB"
/feature="putative"
/number=2
596361..596614
/feature="TRYB"
/feature="putative"
/pseudo
/number=3
596615..597021
/feature="TRYB"
/feature="putative"
/number=3
597022..597158
/feature="TRYB"
/feature="putative"
/pseudo
/number=4
597159..597459
/partial
/feature="TRYB"
/feature="putative"
/number=4
597460..597610
/partial
/feature="TRYB"
/feature="putative"
/pseudo
/number=5
597643..597648
/feature="TRYB"
/feature="putative"
complement(599961..600205)
/feature="putative"
/rpt_family="Alu"
complement(600520..600879)
/feature="putative"
/rpt_family="LINE 1"
Join(604575..604698,605728..605887,606946..607199,
607595..607731,608031..608181)

```

```

/partial
/feature="TRYC"
/feature="putative"
604575..604698
/partial
/feature="TRYC"
/feature="putative"
/number=1
Join(604659..604698,605728..605887,606946..607199,
607595..607731,608031..608183)
/feature="TRYC"
/feature="putative; NCBI gi: 624713"
/codon_start=1
/product="trypsinogen C"
/translation="MNPILLILAFVGAAVVFDDDDKIVGGYTCENSVYQVSLMSG
SHFCGGSLISEQWVVSAGHCYKPHIQVRLGEHNIIEVLENEQFINAAKTIIRPKYRI
TLNDIMLIKLSPTPAVINAHVSTISLPTAPPAAGTECLISGNGTLLSSGADYDELQC
LDAPVLTQAKCKASYPLKITSKMF CVGFLEGGKDCQCGSGGPPVQNCGLAQIVSMGY
GCAQRRPGVYTRVYVYVDMIKDTAANS"
604699..605727
/feature="TRYC"
/feature="putative"
/number=1
605510..647568
/feature=" (vector pWE15A) "
/organism="Homo sapiens"
/cell_line="C6W1"
/germline
/sequenced_mol="DNA"
/clone_lib="Kai Wang's"
/citation=[1]
605728..605887
/feature="TRYC"
/feature="putative"
/number=2
605888..606945
/feature="TRYC"
/feature="putative"
/number=2
606946..607199
/feature="TRYC"
/feature="putative"
/number=3
607200..607594
/feature="TRYC"
/feature="putative"
/number=3
607595..607731
/feature="TRYC"
/feature="putative"
/number=4
607732..608030
/feature="TRYC"
/feature="putative"
/number=4
608031..608181
/partial
/feature="TRYC"
/feature="putative"
608214..608219
/feature="TRYC"
/feature="putative"
610541..610782

```

```

/rpt family="Alu"
/repeat_unit complement (611096..611446)
/notes="putative"
/rpt family="LINE 1"
join(614744..614783,615820..615979,617038..617291,
617692..617828,618126..618275)
/genes="TRYD"
/notes="putative"
/pseudo
/product="trypsinogen D"
join(<614744..614783,615820..615979,617038..617291,
617692..617828,618126..618275)
/genes="TRYD"
/notes="putative"
<614744..614783
/genes="TRYD"
/notes="putative"
/pseudo
/number=1
614784..615819
/genes="TRYD"
/notes="putative"
/number=1
615820..615979
/genes="TRYD"
/notes="putative"
/pseudo
/number=2
615980..617037
/genes="TRYD"
/notes="no splice donor; putative; does not fit consensus"
/number=2
617038..617291
/genes="TRYD"
/notes="putative"
/pseudo
/number=3
617292..617691
/genes="TRYD"
/notes="putative"
/number=3
617692..617828
/genes="TRYD"
/notes="putative"
/pseudo
/number=4
617829..618125
/genes="TRYD"
/notes="putative"
/number=4
618126..618275
/genes="TRYD"
/notes="putative"
/pseudo
/number=5
618308..618313
/genes="TRYD"
/notes="putative"
complement(620651..620747)
/notes="putative"
/rpt family="Alu"
complement(620936..621210)
/notes="putative"

```

```

CDS
/rpt family="LINE 1"
join(625027..625066,626095..626254,627312..627565,
627965..628101,628402..628554)
/genes="TRYE"
/notes="putative; NCBI gi: 624714"
/codon_start=1
/product="trypsinogen E"
/translation="MNLILLITFVAANAAPFDDDDKIVGGYICEENSVPYQVSLNSG
YHFCGSLISEQWVVSAGHCTKSRIQVRIGEHNTIELEGNQFINAAKIRHPKYNR
TLNDILLIKLSPAVINSRVSAISLPTAPPAAGTESLISGKNTLSSGADYPDELQC
LDAPVLQAQCEASYPGKITNNMFCVGFLEGKDSQCGSPVVSNGELQCIYSHGY
GCAQKRPGVYTKVYNYVDWIKDTIAANS"
join(<625027..625066,626095..626254,627312..627565,
627965..628101,628402..628554)
/genes="TRYE"
<625027..625066
/genes="TRYE"
/notes="putative"
/number=1
625067..626094
/genes="TRYE"
/notes="putative"
/number=1
626095..626254
/genes="TRYE"
/notes="putative"
/number=2
626255..627311
/genes="TRYE"
/notes="putative"
/number=2
627312..627565
/genes="TRYE"
/notes="putative"
/number=3
627566..627964
/genes="TRYE"
/notes="putative"
/number=3
627965..628101
/genes="TRYE"
/notes="putative"
/number=4
628102..628401
/genes="TRYE"
/notes="putative"
/number=4
628402..628590
/genes="TRYE"
/notes="putative"
/number=5
628591..628590
/genes="TRYE"
complement(630936..631210)
/notes="putative"
/rpt family="Alu"
complement(631452..631928)
/notes="putative"
/rpt family="LINE 1"
complement(635245..635510)
/notes="putative"
/rpt family="Alu"
complement(635729..635993)

```

```

repeat_unit /note="putative"
            /rpt_family="Alu"
            635948..636654
repeat_unit /note="putative"
            /rpt_family="HOMERSP1B"
            complement(636726..636988)
repeat_unit /note="putative"
            /rpt_family="Alu"
            complement(637950..638214)
misc_signal /note="putative"
            /rpt_family="Alu"
            640240..640267
            /gene="TCRBD1"
misc_signal /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            640240..640248
            /gene="TCRBD1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640249..640260
            /gene="TCRBD1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640261..640267
            /gene="TCRBD1"
D_segment /note="RSS heptamer - awaiting approval of new feature
            key"
            640268..640279
            /gene="TCRBD1"
misc_signal /note="putative"
            640280..640318
            /gene="TCRBD1"
misc_signal /note="RSS - awaiting approval of new feature key;
            heptamer is at 5' end; putative"
            640280..640286
            /gene="TCRBD1"
misc_signal /note="RSS heptamer - awaiting approval of new feature
            key"
            640287..640309
            /gene="TCRBD1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640310..640318
            /gene="TCRBD1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640307..640334
            /gene="TCRBJIS1"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            640307..640315
            /gene="TCRBJIS1"
misc_signal /note="RSS nonamer - awaiting approval of new feature key"
            640316..640327
            /gene="TCRBJIS1"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            640328..640334
            /gene="TCRBJIS1"
            /note="RSS heptamer - awaiting approval of new feature
            key"
            640335..640382
            /gene="TCRBJIS1"
            /note="putative"
            641044..641071
            /gene="TCRBJIS2"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            641044..641052

```

```

misc_signal /gene="TCRBJIS2"
            /note="RSS nonamer - awaiting approval of new feature key"
            641053..641064
            /gene="TCRBJIS2"
misc_signal /note="RSS spacer - awaiting approval of new feature key"
            641065..641071
            /gene="TCRBJIS2"
            /note="RSS heptamer - awaiting approval of new feature
            key"
            641072..641119
            /gene="TCRBJIS2"
            /note="putative"
            641657..641684
            /gene="TCRBJIS3"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            641657..641665
            /gene="TCRBJIS3"
            /note="RSS nonamer - awaiting approval of new feature key;
            putative"
            641666..641677
            /gene="TCRBJIS3"
            /note="RSS spacer - awaiting approval of new feature key"
            641678..641684
            /gene="TCRBJIS3"
            /note="RSS heptamer - awaiting approval of new feature
            key; putative"
            641685..641734
            /gene="TCRBJIS3"
            /note="putative"
            642252..642279
            /gene="TCRBJIS4"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            642252..642260
            /gene="TCRBJIS4"
            /note="RSS nonamer - awaiting approval of new feature key"
            642261..642272
            /gene="TCRBJIS4"
            /note="RSS spacer - awaiting approval of new feature key"
            642273..642279
            /gene="TCRBJIS4"
            /note="RSS heptamer - awaiting approval of new feature
            key"
            642280..642330
            /gene="TCRBJIS4"
            /note="putative"
            642525..642552
            /gene="TCRBJIS5"
            /note="RSS - awaiting approval of new feature key; nonamer
            is at 5' end; putative"
            642525..642533
            /gene="TCRBJIS5"
            /note="RSS nonamer - awaiting approval of new feature key"
            642534..642545
            /gene="TCRBJIS5"
            /note="RSS spacer - awaiting approval of new feature key"
            642546..642552
            /gene="TCRBJIS5"
            /note="RSS heptamer - awaiting approval of new feature
            key"
            642553..642602
            /gene="TCRBJIS5"
            /note="putative"

```

```

misc_signal 643015..643042
/gene="TCRBJ1S6"
/notes="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
643015..643023
/gene="TCRBJ1S6"
/notes="RSS nonamer - awaiting approval of new feature key"
643024..643035
/gene="TCRBJ1S6"
/notes="RSS spacer - awaiting approval of new feature key"
643036..643042
/gene="TCRBJ1S6"
/notes="RSS heptamer - awaiting approval of new feature
key"
J_segment 643043..643095
/gene="TCRBJ1S6"
/notes="putative"
repeat_unit complement(643863..644120)
/notes="putative"
/rpt_family="Alu"
CDS join(645749..646135,646577..646594,646747..646853,
647176..647196)
/partial
/gene="TCRBC1"
/notes="This CDS feature is included to show the
translation of the corresponding C region. Presently
translation qualifiers on C_region features are illegal.
NCBI gi: 624715"
/codon_start=3
/translating="DINKVPPPEVAFEPSEAFISHTQKATLVCIATCFEPDHVLSW
WNGKGVHSGVSTDPQPIKEQPALNDSRYCLSSRLVSAFTWONPRHFRQVQFYGL
SENDEWTDRAKPYVTQIVSAEAWGRADCGFTSVSYQQCVLSATILYEILIGKATLYAV
LVSAVLVMAWKRKDF"
C_region join(645749..646135,646577..646594,646747..646853,
647176..647196)
/gene="TCRBC1"
/notes="putative"
/codon_start=3
645749..646135
/gene="TCRBC1"
/notes="putative"
/number=1
646136..646576
/gene="TCRBC1"
/notes="putative"
/number=1
646577..646594
/gene="TCRBC1"
/notes="putative"
/number=2
646595..646746
/gene="TCRBC1"
/notes="putative"
/number=2
646747..646853
/gene="TCRBC1"
/notes="putative"
/number=3
646854..647175
/gene="TCRBC1"
/notes="putative"
/number=3
647176..647403
/gene="TCRBC1"

```

```

/notes="putative"
/number=4
647326..650397
/notes="The gap between cosmid clones CBG1 and C29 was
closed by the sequencing of a PCR reaction product."
/organism="Homo sapiens"
/germline
/sequenced_mol="DNA"
/clone="CBG1/C29gap"
/citation=[1]
647398..647403
/gene="TCRBC1"
/notes="putative"
649732..649759
/gene="TCRBD2"
/notes="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
649732..649740
/gene="TCRBD2"
/notes="RSS nonamer - awaiting approval of new feature key"
649741..649752
/gene="TCRBD2"
/notes="RSS spacer - awaiting approval of new feature key"
649753..649759
/gene="TCRBD2"
/notes="RSS heptamer - awaiting approval of new feature
key"
649760..649771
/gene="TCRBD2"
/notes="putative"
649776..649814
/gene="TCRBD2"
/notes="RSS - awaiting approval of new feature key;
heptamer is at 5' end; putative"
649776..649782
/gene="TCRBD2"
/notes="RSS heptamer - awaiting approval of new feature
key"
649783..649805
/gene="TCRBD2"
/notes="RSS spacer - awaiting approval of new feature key"
649806..649814
/gene="TCRBD2"
/notes="RSS nonamer - awaiting approval of new feature key"
650285..687973
/notes="(vector pWE15A)"
/organism="Homo sapiens"
/cell_line="CCM1"
/germline
/sequenced_mol="DNA"
/clone="C29"
/clone_lib="Kai Wang's"
/citation=[1]
650392..650419
/gene="TCRBJ2S1"
/notes="RSS - awaiting approval of new feature key; nonamer
is at 5' end; putative"
650392..650400
/gene="TCRBJ2S1"
/notes="RSS nonamer - awaiting approval of new feature key"
650401..650412
/gene="TCRBJ2S1"
/notes="RSS spacer - awaiting approval of new feature key"
650413..650419

```



```

/codon_start=3
exon      655095..655481
          /gene="TCRBC2"
          /note="putative"
          /number=1
intron    655482..655997
          /gene="TCRBC2"
          /note="putative"
          /number=1
exon      655998..656015
          /gene="TCRBC2"
          /note="putative"
          /number=2
intron    656016..656158
          /gene="TCRBC2"
          /note="putative"
          /number=2
exon      656159..656265
          /gene="TCRBC2"
          /note="putative"
          /number=3
intron    656266..656556
          /gene="TCRBC2"
          /note="putative"
          /number=3
exon      656557..656774
          /gene="TCRBC2"
          /note="putative"
          /number=4
polyA_signal
          656769..656774
          /gene="TCRBC2"
          /note="putative"
          complement(658014..658405)
repeat_unit
          /note="putative"
          /rpt_family="MLT1B"
repeat_unit
          complement(665732..665995)
          /note="putative"
          /rpt_family="Alu"
misc_signal
          complement(666601..666639)
          /gene="TCRBC20S1"
          /note="RSS - awaiting approval of new feature key; nonamer
          is at 5' end; putative"
misc_signal
          666601..666609
          /note="RSS_nonamer - awaiting approval of new feature key"
misc_signal
          666610..666632
          /note="RSS_spacer - awaiting approval of new feature key"
misc_signal
          666633..666639
          /note="RSS_heptamer - awaiting approval of new feature
          key"
CDS
          complement(join(666640..666931,667298..667340))
          /partial
          /gene="TCRBCV20S1"
          /note="This CDS feature is included to show the
          translation of the corresponding V segment. Presently
          translation qualifiers on V_segment features are illegal.
          NCBI gi: 624717"
          /codon_start=1
          /translation="MGLSLALLCTFFGVRSGTHQWPATIVQVGPISLCTVEG
          TSNPNLWYRQAGRGQLLFYSVGIQGISSEVPQNLASRPDRQFILSKKLLISD
          SGFYICAWSV"
          complement(join(666640..666931,667298..667340))
          /gene="TCRBCV20S1"
          /note="putative"
          /codon_start=1

```

```

exon      complement(666640..666931)
          /gene="TCRBCV20S1"
          /note="putative"
          /number=2
intron    complement(666932..667297)
          /gene="TCRBCV20S1"
          /note="putative"
          /number=1
exon      complement(667298..>667340)
          /gene="TCRBCV20S1"
          /note="putative"
          /number=1
repeat_unit
          670197..670685
          /note="putative"
          /rpt_family="MLT1D"
repeat_unit
          complement(671437..671825)
          /note="putative"
          /rpt_family="MLT1D"
repeat_unit
          complement(676821..679089)
          /note="putative"
          /rpt_family="LINE 1"
repeat_unit
          680649..680908
          /note="putative"
          /rpt_family="Alu"
repeat_unit
          683515..683762
          /note="putative"
          /rpt_family="Alu"
BASE COUNT 199337 a 146423 c 144088 g 195125 t
ORIGIN
DB 51; Score 26; Match 74.1%; QryMatch 2.4%; Pred. No. 7.09e-03;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Db 322953 cccacattccacatccgcagtcacatccacccccacattccctgccctct 323006
          ||||| ||||| | || ||||| |||| |||||
Qy 931 CCCACCTGCCACCCCTGTGTCCAGCTCCACCCCTGCTTCTGACCCCTCT 984
Search completed: Thu Sep 28 00:16:46 1995
Job time : 624 secs.

```